

(12) NACH DEM VERTRAG ÜBER DIE INTERNATIONALE ZUSAMMENARBEIT AUF DEM GEBIET DES  
PATENTWESENS (PCT) VERÖFFENTLICHTE INTERNATIONALE ANMELDUNG

(19) Weltorganisation für geistiges Eigentum  
Internationales Büro



(43) Internationales Veröffentlichungsdatum  
14. Juni 2001 (14.06.2001)

PCT

(10) Internationale Veröffentlichungsnummer  
**WO 01/42493 A2**

- (51) Internationale Patentklassifikation?: C12Q 1/68 (81) Bestimmungsstaaten (*national*): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.
- (21) Internationales Aktenzeichen: PCT/DE00/04381 (84) Bestimmungsstaaten (*regional*): ARIPO-Patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), eurasisches Patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), europäisches Patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI-Patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).
- (22) Internationales Anmeldedatum:  
6. Dezember 2000 (06.12.2000)
- (25) Einreichungssprache: Deutsch
- (26) Veröffentlichungssprache: Deutsch
- (30) Angaben zur Priorität:  
199 59 691.3 6. Dezember 1999 (06.12.1999) DE
- (71) Anmelder (*für alle Bestimmungsstaaten mit Ausnahme von US*): EPIGENOMICS AG [DE/DE]; Kastanienallee 24, D-10435 Berlin (DE).
- (72) Erfinder; und
- (75) Erfinder/Anmelder (*nur für US*): OLEK, Alexander [DE/DE]; Kyffhäuserstrasse 20, D-10781 Berlin (DE). PIEPENBROCK, Christian [DE/DE]; Schwartzkopffstrasse 7 B, D-10115 Berlin (DE).
- (74) Anwalt: SCHUBERT, Clemens; Joachimstrasse 9, D-10119 Berlin (DE).

Veröffentlicht:

— Ohne internationalem Recherchenbericht und erneut zu veröffentlichen nach Erhalt des Berichts.

Zur Erklärung der Zweibuchstaben-Codes, und der anderen Abkürzungen wird auf die Erklärungen ("Guidance Notes on Codes and Abbreviations") am Anfang jeder regulären Ausgabe der PCT-Gazette verwiesen.

(54) Title: METHOD FOR THE PARALLEL DETECTION OF THE DEGREE OF METHYLATION OF GENOMIC DNA

(54) Bezeichnung: VERFAHREN ZUR PARALLELEN DETEKTION DES METHYLIERUNGZUSTANDES VON GENOMISCHER DNA

(57) Abstract: The invention relates to a method for the parallel detection of the degree of methylation of genomic DNA wherein the following steps are performed: (a) chemical treatment at the 5' position of non-methylated cytosine bases converts said bases into uracil, thymidine or another base which exhibits hybridization behavior different to that of cytosine in a genomic DNA sample; (b) more than ten different fragments, each having less than 2000 base pairs in said chemically treated genomic DNA sample, are amplified simultaneously using synthetic oligonucleotides as a primer, whereby said primers each contain genomic sequences which are involved in gene regulation and/or transcribed and/or translated, such as those sequences which should be obtained after execution of steps (a); (c) the sequence contexts of all or a portion of the CpG dinucleotides or CpNpG trinucleotides contained in the amplified fragments are determined.

**WO 01/42493 A2**

(57) Zusammenfassung: Beschrieben ist ein Verfahren zur parallelen Detektion des Methylierungszustandes von genomicscher DNA bei dem man folgende Schritte ausführt: (a) in einer genomicschen DNA Probe wandelt man durch chemische Behandlung an der 5'-Position unmethylierte Cytosinbasen in Uracil, Thymidin oder eine andere vom Hybridisierungsverhalten her dem Cytosin unähnliche Base um; (b) aus dieser chemisch behandelten genomicschen DNA amplifiziert man mehr als zehn unterschiedliche Fragmente, die jeweils weniger als 2000 Basenpaare lang sind, gleichzeitig durch Verwendung von synthetischen Oligonukleotiden als Primer, wobei diese Primer jeweils Sequenzen aus an der Genregulation beteiligten und/oder transkribierten und/oder translatierten genomicschen Sequenzen enthalten, wie sie nach einer Behandlung gemäß Schritt (a) vorliegen würden; (c) man bestimmt den Sequenzkontext aller oder eines Teils der in den amplifizierten Fragmenten enthaltenen CpG Dinukleotide oder CpNpG Trinukleotide.

**Verfahren zur parallelen Detektion des Methylierungszustandes  
von genomischer DNA**

Die vorliegende Erfindung betrifft ein Verfahren zur parallelen Detektion des Methylierungszustandes von genomischer DNA.

Die nach den methodischen Entwicklungen der letzten Jahre in der Molekularbiologie gut studierten Beobachtungsebenen sind die Gene selbst, die Übersetzung dieser Gene in RNA und die daraus entstehenden Proteine. Wann im Laufe der Entwicklung eines Individuums welches Gen angeschaltet wird und wie Aktivieren und Inhibieren bestimmter Gene in bestimmten Zellen und Geweben gesteuert wird, ist mit hoher Wahrscheinlichkeit mit Ausmaß und Charakter der Methylierung der Gene bzw. des Genoms korrelierbar. Insofern ist die Annahme naheliegend, daß pathogene Zustände sich in einem veränderten Methylierungsmuster einzelner Gene oder des Genoms äußern.

Stand der Technik sind Verfahren, welche das Studium von Methylierungsmustern einzelner Gene gestatten. Jüngere Fortentwicklungen dieser Methode erlauben auch die Analyse kleinsten Mengen an Ausgangsmaterial. Die vorliegende Erfindung beschreibt ein Verfahren zur parallelen Detektion des Methylierungszustandes genomicscher DNA Proben, wobei ausgehend von einer Probe gleichzeitig zahlreiche verschiedenene Fragmente aus an der Genregulation beteiligten oder/und transkribierten und/oder translatierten Sequenzen amplifiziert werden und anschließend der Sequenzkontext in den amplifizierten Fragmenten enthaltenen CpG Dinukleotide untersucht wird.

5-Methylcytosin ist die häufigste kovalent modifizierte Base in der DNA eukaryotischer Zellen. Sie spielt beispielsweise eine Rolle in der Regulation der Transkription, genomischem Imprinting und in der Tumorgenese. Die Identifizierung von 5-Methylcytosin als Bestandteil genetischer Information ist daher von erheblichem Interesse. 5-Methylcytosin-Positionen können jedoch nicht durch Sequenzierung identifiziert werden, da 5-Methylcytosin das gleiche Basenpaarungsverhalten aufweist

wie Cytosin. Darüber hinaus geht bei einer PCR-Amplifikation die epigenetische Information, welche die 5-Methylcytosine tragen, vollständig verloren.

Die Modifikation der genomischen Base Cytosin zu 5'-Methylcytosin stellt den bis heute wichtigsten und best untersuchten epigenetischen Parameter dar. Trotzdem gibt es bis heute zwar Methoden, umfassende Genotypen von Zellen und Individuen zu ermitteln, aber noch keine vergleichbaren Ansätze auch in großem Maße epigenotypische Information zu generieren und auszuwerten.

Es sind im Prinzip drei prinzipiell verschiedene Methoden bekannt, den 5-Methyl-Status eines Cytosins im Sequenzkontext zu bestimmen.

Die erste prinzipielle Methode beruht auf der Verwendung von Restriktionsendonukleasen (RE), welche „methylierungssensitiv“ sind. REs zeichnen sich dadurch aus, daß sie an einer bestimmten DNA-Sequenz, meist zwischen 4 und 8 Basen lang, einen Schnitt in die DNA einführen. Die Position solcher Schnitte kann dann durch Gelektrophorese, Transfer auf eine Membran und Hybridisierung nachgewiesen werden. Methylierungssensitiv bedeutet, daß bestimmte Basen innerhalb der Erkennungssequenz unmethyliert vorliegen müssen, damit der Schnitt erfolgen kann. Das Bandenmuster nach einem Restriktionsschnitt und Gelektrophorese ändert sich also je nach Methylierungsmuster der DNA. Allerdings befinden sich die wenigsten methylierbaren CpG innerhalb von Erkennungssequenzen von REs, können also nach dieser Methode nicht untersucht werden.

Die Empfindlichkeit dieser Methoden ist extrem niedrig (Bird, A.P., and Southern, E.M., J.Mol. Biol. 118, 27-47). Eine Variante kombiniert PCR mit dieser Methode, eine Amplifikation durch zwei auf beiden Seiten der Erkennungssequenz liegende Primer erfolgt nach einem Schnitt nur dann, wenn die Erkennungssequenz methyliert vorliegt. Die Empfindlichkeit steigt in diesem Fall auf theoretisch ein einziges Molekül der Zielsequenz, allerdings können mit hohem Aufwand nur einzelne Positionen untersucht werden (Shemer, R. et al., PNAS 93, 6371-6376). Wiederum ist Voraussetzung, daß sich die methylierbare Position innerhalb der Erkennungssequenz einer RE befindet.

Die zweite Variante beruht auf partieller chemischer Spaltung von Gesamt-DNA, nach dem Vorbild einer Maxam-Gilbert Sequenziereaktion, Ligation von Adaptoren an die so generierten Enden, Amplifikation mit generischen Primern und Auf trennung auf einer Gelektrophorese. Mit diesem Verfahren können definierte Bereiche bis zur Größe von weniger als tausend Basenpaaren untersucht werden. Das Verfahren ist allerdings so kompliziert und unzuverlässig, daß es praktisch nicht mehr verwendet wird (Ward, C. et al., J. Biol. Chem. 265, 3030-3033).

Eine relativ neue und die mittlerweile am häufigsten angewandte Methode zur Untersuchung von DNA auf 5-Methylcytosin beruht auf der spezifischen Reaktion von Bisulphit mit Cytosin, das nach anschließender alkalischer Hydrolyse in Uracil umgewandelt wird, welches in seinem Basen-Paarungsverhalten dem Thymidin entspricht. 5-Methylcytosin wird dagegen unter diesen Bedingungen nicht modifiziert. Damit wird die ursprüngliche DNA so umgewandelt, daß Methylcytosin, welches ursprünglich durch sein Hybridisierungsverhalten vom Cytosin nicht unterschieden werden kann, jetzt durch „normale“ molekularbiologische Techniken als einzig verbliebenes Cytosin beispielsweise durch Amplifikation und Hybridisierung oder Sequenzierung nachgewiesen werden kann. Alle diese Techniken beruhen auf Basenpaarung, welche jetzt voll ausgenutzt werden kann. Der Stand der Technik, was die Empfindlichkeit betrifft, wird durch ein Verfahren definiert, welches die zu untersuchende DNA in einer Agarose-Matrix einschließt, dadurch die Diffusion und Renaturierung der DNA (Bisulphit reagiert nur an einzelsträngiger DNA) verhindert und alle Fällungs- und Reinigungsschritte durch schnelle Dialyse ersetzt (Olek, A. et al., Nucl. Acids. Res. 24, 5064-5066). Mit dieser Methode können einzelne Zellen untersucht werden, was das Potential der Methode veranschaulicht. Allerdings werden bisher nur einzelne Regionen bis etwa 3000 Basenpaare Länge untersucht, eine globale Untersuchung von Zellen auf Tausende von möglichen Methylierungereignissen ist nicht möglich. Allerdings kann auch dieses Verfahren keine sehr kleinen Fragmente aus geringen Probenmengen zuverlässig analysieren. Diese gehen trotz Diffusionsschutz durch die Matrix verloren.

Eine Übersicht über die weiteren bekannte Möglichkeiten, 5-Methylcytosine

nachzuweisen, kann auch dem folgenden Übersichtsartikel entnommen werden: Rein, T., DePamphilis, M. L., Zorbas, H., Nucleic Acids Res. 26, 2255 (1998).

Die Bisulphit-Technik wird bisher bis auf wenige Ausnahmen (z.B. Zeschnigk, M. et al., Eur. J. Hum. Gen. 5, 94-98; Kubota T. et al., Nat. Genet. 16, 16-17) nur in der Forschung angewendet. Immer aber werden kurze, spezifische Stücke eines bekannten Gens nach einer Bisulphit-Behandlung amplifiziert und entweder komplett sequenziert (Olek, A. and Walter, J., Nat. Genet. 17, 275-276) oder einzelne Cytosin-Positionen durch eine „Primer-Extension-Reaktion“ (Gonzalgo, M. L. and Jones, P.A., Nucl. Acids. Res. 25, 2529-2531) oder Enzymschnitt (Xiong, Z. and Laird, P.W., Nucl. Acids. Res. 25, 2532-2534) nachgewiesen. Zudem ist auch der Nachweis durch Hybridisierung beschrieben worden (Olek et al, WO9928498).

Gemeinsamkeiten zwischen Promotoren bestehen nicht nur im Vorkommen von TATA- oder GC-Boxen sondern auch darin, für welche Transkriptionsfaktoren sie Bindestellen besitzen und in welchem Abstand diese sich zueinander befinden. Die existierenden Bindestellen für ein bestimmtes Protein stimmen in ihrer Sequenz nicht vollständig überein, es finden sich aber konservierte Folgen von mindestens 4 Basen, die durch das Einfügen von „Wobbles“, d. h. Positionen, an denen sich jeweils unterschiedliche Basen befinden, noch verlängert werden können. Des weiteren liegen diese Bindestellen in bestimmten Abständen zueinander vor.

Die Verteilung der DNA im Interphase-Chromatin, das den größten Teil des nuklearen Volumens einnimmt, unterliegt jedoch einer ganz speziellen Ordnung. So ist die DNA an mehreren Stellen an die nukleare Matrix, eine filamentöse Struktur an der Innenseite der nuklearen Membran, angeheftet. Diese Regionen, bezeichnet man als matrix attachment regions (MAR) oder scaffold attachment regions (SAR). Das Anheften hat wesentlichen Einfluß auf die Transkription bzw. die Replikation. Diese MAR-Fragmente weisen keine konservativen Sequenzen auf, bestehen allerdings zu 70% aus A bzw. T und liegen in der Nähe von cis-agierenden Regionen, die die Transkription allgemein regulieren, und Topoisomerase II-Erkennungsstellen.

Neben Promotoren und Enhancern existieren weitere regulatorische Elemente für verschiedene Gene, sogenannte Insulators. Diese Insulators können z.B. die Wirkung des Enhancers auf den Promotor inhibieren, wenn sie zwischen Enhancer und Promotor liegen, oder aber, zwischen Heterochromatin und einem Gen gelegen, das aktive Gen vor dem Einfluß des Heterochromatins schützen. Beispiele für solche Insulators sind: 1. sogenannte LCR (locus control regions), welche aus mehreren gegenüber DNAase I hypersensitiven Stellen besteht; 2. bestimmte Sequenzen wie SCS (specialized chromatin structures) bzw. SCS', 350 bzw. 200 bp lang und hochresistent gegen Degradierung durch DNAase I und auf beiden Seiten von hypersensitiven Stellen flankiert (Abstand je 100 bp). An scs' bindet das Protein BEAF-32. Diese Insulators können auf beiden Seiten des Gens liegen.

Eine Übersicht über den Stand der Technik in der Oligomer Array Herstellung läßt sich auch einer im Januar 1999 erschienenen Sonderausgabe von Nature Genetics (Nature Genetics Supplement, Volume 21, January 1999) und der dort zitierten Literatur entnehmen.

Patente, die sich allgemein auf die Verwendung von Oligomer Arrays und photolithographisches Maskendesign beziehen, sind z. B. US-A 5,837,832, US-A 5,856,174, WO-A 98/27430 und US-A 5,856,101. Zudem existieren einige Stoff- und Verfahrenspatente, welche die Verwendung photolabiler Schutzgruppen an Nukleosiden einschränken, so z. B. WO-A98/39348 und US-A 5,763,599.

Matrix-assistierte Laser Desorptions/Ionisations Massenspektrometrie (MALDI) ist eine neue, sehr leistungsfähige Entwicklung für die Analyse von Biomolekülen (Karas, M. and Hillenkamp, F. 1988. Laser desorption ionization of proteins with molecular masses exceeding 10.000 daltons. Anal. Chem. 60: 2299-2301). Ein Analytmolekül wird in eine im UV absorbierende Matrix eingebettet. Durch einen kurzen Laserpuls wird die Matrix ins Vakuum verdampft und das Analyt so unfragmentiert in die Gasphase befördert. Eine angelegte Spannung beschleunigt die Ionen in ein feldfreies Flugrohr. Auf Grund ihrer verschiedenen Massen werden Ionen unterschiedlich stark beschleunigt. Kleinere Ionen erreichen den Detektor früher als größere und die Flugzeit wird in die Masse der

Ionen umgerechnet.

Für die Abtastung eines immobilisierten DNA-Arrays sind vielfach fluoreszent markierte Sonden verwendet worden. Besonders geeignet sind für die Fluoreszenzmarkierung ist das einfache Anbringen von Cy3 und Cy5 Farbstoffen am 5'OH der jeweiligen Sonde. Die Detektion der Fluoreszenz der hybridisierten Sonden erfolgt beispielsweise über ein Konfokalmikroskop. Die Farbstoffe Cy3 und Cy5 sind, neben vielen anderen, kommerziell erhältlich.

Um die erwartete Anzahl von amplifizierten Fragmenten ausgehend von einer beliebigen Templat-DNA und zweien nicht für jeweils eine bestimmte Position spezifischen Primern zu berechnen, muß ein statistisches Modell über den Aufbau des Genoms zu Grunde gelegt werden.

Wir geben hier die Berechnung für drei Modelle an, beziehen uns allerdings in diesem Patent auf die in Modell 3 beschriebene Methode.

Modell 1:

Im einfachsten Fall wird angenommen, daß ein primärer DNA-Strang eine Zufallsfolge von vier gleich häufig vorkommenden Basen ist. Damit ergibt sich als Wahrscheinlichkeit, daß sich für einen beliebigen Primer *PrimA* (der Länge  $k$ ) an einer gegebenen Stelle im Genom eine perfekte Basenpaarung ergibt:

$$P_a(\text{PrimA}) = 0.25^k \quad (\text{Modell 1 für DNA})$$

(diese Wahrscheinlichkeit ist für den sense- und anti-sense-Strang der DNA gleich)

Bei einer Bisulfitbehandlung der DNA werden diejenigen Cytosine durch Uracil ersetzt, die nicht zu einem methylierten CG gehören. Das Basenpaarungsverhalten des Uracils entspricht dem des Thymins. Da CG in der DNA sehr selten sind (unter zwei Prozent), kann die statistische Häufigkeit der Cs nach der Bisulfitbehandlung vernachlässigt werden. Die Wahrscheinlichkeit, daß sich für einen Primer *PrimB* (Länge  $k$ , davon  $a$  As,

*t* Ts, *g* Gs und *c* Cs) auf bisulfitbehandelter DNA eine perfekte Basenpaarung ergibt, ist unterschiedlich für einen mit Bisulfit behandelten Strang und den zugehörigen anti-sense Strang:

$$P_{Is}(PrimB)=0.5^a * 0.25^t * 0.25^c * 0^g \quad (\text{Modell 1 für Bisulfit-DNA-Strang})$$

$$P_{Ia}(PrimB)=0.25^a * 0.5^t * 0^c * 0.25^g \quad (\text{Modell 1 für anti-sense-Strang zu einem Bisulfit-DNA-Strang})$$

(wenn der Primer C oder G enthält, wird somit einer der Wahrscheinlichkeitswerte 0).

### Modell 2:

Zählungen der Basenhäufigkeiten der DNA ergeben, daß die vier Basen in der DNA nicht gleichverteilt sind. Entsprechend kann man aus DNA-Datenbanken folgende Häufigkeiten (Wahrscheinlichkeiten für ein Vorkommen) der Basen ermitteln.

$$P_{DNA}(A)=0.2811$$

$$P_{DNA}(T)=0.2784$$

$$P_{DNA}(C)=0.2206$$

$$P_{DNA}(G)=0.2199$$

Als Grundlage für diese Statistik (und die folgenden für Modell 2 und 3) dienen ca. 6% des Genoms vom Homo Sapiens aus High Throughput Sequencing Projekten (Datenbank "htgs" vom NIH/NCBI vom 6.9.1999). Die Gesamtmenge der Daten beträgt mehr als  $1.5 \times 10^8$  Basenpaare, was einem Schätzfehler für die Einzelwahrscheinlichkeiten kleiner  $10^{-5}$  entspricht.

Mit Hilfe dieser Werte lässt sich das Modell 1 verbessern.

Damit ist die Wahrscheinlichkeit, daß sich für einen Primer *PrimC* (Länge *k*, davon *a* As, *t* Ts, *g* Gs und *c* Cs) eine perfekte Basenpaarung ergibt:

$$P_2(PrimC) = P_{DNA}(T)^a * P_{DNA}(A)' * P_{DNA}(C)^g * P_{DNA}(G)^c \quad (\text{Modell 3 für DNA})$$

Für den mit Bisulfit behandelten Strang ergeben sich folgende Wahrscheinlichkeiten unter der Annahme, daß alle CpG-Positionen methyliert sind (man erhält eine gleiche Statistik für die Bisulfitbehandlung des DNA-sense- und des DNA-antisense-Stranges):

$$P_{bDNA}(A) = 0.2811$$

$$P_{bDNA}(C) = 0.0140$$

$$P_{bDNA}(G) = 0.2199$$

$$P_{bDNA}(T) = 0.4850$$

Damit ergibt sich als Wahrscheinlichkeit, daß sich für einen Primer *PrimD* (Länge  $k$ , davon  $a$  As,  $t$  Ts,  $g$  Gs und  $c$  Cs) eine perfekte Basenpaarung ergibt:

$$P_{2s}(PrimD) = P_{bDNA}(T)^a * P_{bDNA}(A)' * P_{bDNA}(C)^g * P_{bDNA}(G)^c \quad (\text{Modell 3 für Bisulfit-DNA-Strang})$$

$$P_{2a}(PrimD) = P_{bDNA}(A)^a * P_{bDNA}(T)' * P_{bDNA}(G)^g * P_{bDNA}(C)^c \quad (\text{Modell 3 für anti-sense-Strang zu einem Bisulfit-DNA-Strang})$$

### Modell 3:

Wesentliche Schätzfehler in Modell 2 ergeben sich vor allem bei der mit Bisulfit behandelten DNA aus der Tatsache, daß C nur noch im Kontext CG auftreten kann. Modell 3 berücksichtigt diese Eigenschaft und nimmt an, daß die primäre DNA eine Zufallsfolge mit Abhängigkeit direkt benachbarter Basen ist (Markov-Kette erster Ordnung). Die empirisch aus der Datenbank (vollständig methyliert; mit Bisulfat behandelt) ermittelten paarweisen Basenwahrscheinlichkeiten ergeben sich gleich für beide DNA-Stränge als  $P_{bDNA}(\text{von};\text{nach})$  aus der folgenden Tabelle:

Von\nach	A	C	G	T
A	0.0894	0.0033	0.0722	0.1162
C	0.0	0.0	0.0140	0.0
G	0.0603	0.0036	0.0601	0.0959
T	0.1314	0.0071	0.0736	0.2729

$$P_{bDNA}(A) = 0.2811$$

$$P_{bDNA}(C) = 0.0140$$

$$P_{bDNA}(G) = 0.2199$$

$$P_{bDNA}(T) = 0.4850$$

und für den dazu revers-komplementären Strang (durch entsprechendes Austauschen der Einträge)  $P_{rbDNA}(\text{von ; nach})$

Von\nach	A	C	G	T
A	0.2729	0.0959	0.0	0.1162
C	0.0736	0.0601	0.0140	0.0722
G	0.0071	0.0036	0.0	0.0033
T	0.1314	0.0603	0.0	0.0894

$$P_{rbDNA}(A) = 0.4850 ;$$

$$P_{rbDNA}(C) = 0.2199$$

$$P_{rbDNA}(G) = 0.0140$$

$$P_{rbDNA}(T) = 0.2811$$

Damit hängt die Wahrscheinlichkeit, daß sich für einen Primer *PrimE* (mit der Basenfolge  $B_1 B_2 B_3 B_4 \dots$ ; z.B. ATTG...) eine perfekte Basenpaarung ergibt, von der genauen Abfolge der Basen ab und ergibt sich als das Produkt:

$$P_{3s}(\text{PrimE}) = P_{rbDNA}(B_1) \frac{P_{rbDNA}(B_1; B_2)}{P_{rbDNA}(B_1)} \frac{P_{rbDNA}(B_2; B_3)}{P_{rbDNA}(B_2)} \frac{P_{rbDNA}(B_3; B_4)}{P_{rbDNA}(B_3)} \dots \quad (\text{Modell 3 für})$$

Bisulfit-DNA-Strang)

$$P_{3a}(PrimE) = P_{bDNA}(B_1) \frac{P_{bDNA}(B_1; B_2)}{P_{bDNA}(B_1)} \frac{P_{bDNA}(B_2; B_3)}{P_{bDNA}(B_2)} \frac{P_{bDNA}(B_3; B_4)}{P_{bDNA}(B_3)} \dots \quad (\text{Modell 3 für anti-sense-Strang zu einem Bisulfit-DNA-Strang})$$

Berechnung der Anzahl der zu erwartenden amplifizierten Fragmente

Die mit Bisulfit behandelte DNA wird unter Benutzung einer Anzahl Primer amplifiziert. Aus Sicht des Modells besteht die DNA aus je einem sense- und einem anti-sense-Strang der Länge  $N$  Basen (alle Chromosomen werden hier zusammengefaßt). Für einen Primer  $Prim$  ist zu erwarten, daß er auf dem sense-Strang

$$N * P_s(Prim)$$

perfekte Basenpaarungen ergibt - für diese Berechnung können die Funktionen  $P_{1s}$ ,  $P_{2s}$  oder  $P_{3s}$  von Modell 1, 2 oder 3 eingesetzt werden, je nach gewünschter Abschätzungsgüte. Werden mehrere Primer ( $PrimU$ ,  $PrimV$ ,  $PrimW$ ,  $PrimX$ , etc.) gleichzeitig verwendet, ergibt sich als Wahrscheinlichkeit für eine perfekte Basenpaarung auf dem sense-Strang an einer gegebenen Position:

$$\begin{aligned} P_s(Primers) &= P_s(PrimU) \\ &+ (1 - P_s(PrimU))P_s(PrimV) \\ &+ (1 - P_s(PrimU))(1 - P_s(PrimV))P_s(PrimW) \\ &+ (1 - P_s(PrimU))(1 - P_s(PrimV))(1 - P_s(PrimW))P_s(PrimX) \\ &+ \dots \end{aligned}$$

Und damit als Anzahl der zu erwartenden perfekten Basenparungen mit irgendeinem der Primer

$$N * P_s(Primers)$$

Für die Bestimmung von  $P_a(Primers)$  auf dem anti-sense-Strang werden die analogen Gleichungen verwendet. Ein Amplifikat entsteht genau dann, wenn bei einer perfekten Basenpaarung auf dem sense-Strang innerhalb der maximalen Fragmentlänge  $M$  ein Primer auf dem Gegenstrang eine perfekte Basenpaarung bildet. Die Wahrscheinlichkeit dafür ist

$$P_a(Primers) \sum_{i=0}^{M-2} (1 - P_a(Primers))$$

Für große  $M$  und kleine  $P_a(Primers)$  kann dieses durch folgenden Ausdruck berechnet werden:

$$\frac{1 - P_a(Primers)}{\log(1 - P_a(Primers))} [(1 - P_a(Primers))^M - 1]$$

Für die Gesamtzahl  $F$  der Fragmente, die durch die Amplifikation beider Stränge zu erwarten sind, ergibt sich damit:

$$F = N * P_s(Primers) \frac{(1 - P_a(Primers))}{\log(1 - P_a(Primers))} [(1 - P_a(Primers))^M - 1] + N * P_a(Primers) \frac{(1 - P_s(Primers))}{\log(1 - P_s(Primers))} [(1 - P_s(Primers))^M - 1]$$

Diese Methode liefert einen präzisen Erwartungswert für die Vorhersage der Anzahl der Bindungssites bestimmter Sequenzen an ein beliebiges zuvor mit Bisulfit behandeltes genomicsches DNA Fragment. Sie dient hier als Grundlage für die Berechnung der statistisch erwarteten Anzahl von Amplifikaten in einer PCR-Reaktion ausgehend von zwei Primersequenzen und einer DNA der Länge  $N$ , wobei nur die Amplifikate berücksichtigt werden, die eine Anzahl von  $M$  Nukleotiden nicht überschreiten. In diesem Patent wird davon ausgegangen, daß  $M$  den Wert 2000 hat.

Die bekannten Verfahren für den Nachweis von Cytosin Methylierungen in genomischer DNA sind prinzipiell nicht so ausgelegt, daß eine Vielzahl von Zielregionen im zu untersuchenden Genom gleichzeitig erfaßt werden. Aufgabe der vorliegenden Erfindung ist es, ein Verfahren zu schaffen, mit dem eine Probe genommischer DNA gleichzeitig an mehreren Positionen gleichzeitig auf Cytosin Methylierung hin untersucht werden kann.

Die Aufgabe wird durch die kennzeichnenden Merkmale des Anspruchs 1 gelöst. Vorteilhafte Weiterbildungen der Merkmal sind in den abhängigen Ansprüchen gekennzeichnet.

Im Unterschied zu anderen Verfahren kann nach chemischer Vorbehandlung der DNA durch Verwendung entsprechend angepaßter Primerpaare eine Amplifikation von vielen Zielregionen gleichzeitig erfolgen. Dabei ist es nicht unbedingt notwendig den Sequenzkontext aller dieser Zielregionen vorab zu kennen, da in vielen Fällen, wie nachfolgend auch beispielhaft aufgeführt, Konsensussequenzen aus der Sequenzierung verwandter Zielregionen bekannt sind, die wie unten beschrieben für das Design für bestimmte Zielregionen spezifischer oder selektiver Primerpaare eingesetzt werden können. Das Verfahren ist dann erfolgreich angewandt, wenn die Amplifikation der chemisch vorbehandelten genomischen DNA mehr Fragmente bis maximal 2000 Basenpaare Länge als statistisch zu erwarten aus den jeweils zu untersuchenden Zielregionen liefert.

Dabei wird der statistische Erwartungswert für die Anzahl dieser Fragmente über die im Stand der Technik aufgeführten Formeln berechnet. Die Anzahl der im Amplifikationsschritt hergestellten Fragmente kann dagegen mittels einer beliebigen molekularbiologischen, chemischen oder physikalischen Methode nachgewiesen werden.

Für die Durchführung der erforderlichen statistischen Betrachtungen, die auch für die unten aufgeführten Ansprüche relevant sind, werden die folgenden Werte angenommen:

Das menschliche haploide Genom enthält 3 Milliarden Basenpaare und 100.000 Gene, die wiederum im Mittel eine 2000 Basenpaare lange mRNA codieren, die Gene inklusive der Introns sind durchschnittlich 15000 Basenpaare lang. Promotoren umfassen je Gen 1000 Basenpaare durchschnittlich. Ist daher der statistische Erwartungswert für die Anzahl der Amplifikate, die ausgehend von zwei Primern in transkribierten Sequenzen liegen, zu berechnen, so ist zunächst der Erwartungswert für das Gesamtgenom nach obiger Formel (Methode 3) zu berechnen und mit dem Anteil der transkribierten Sequenzen am Gesamtgenom zu berechnen. Analog wird für Teile eines beliebigen Genoms sowie für Promotoren und translatierte Sequenzen (mRNA codierend) vorgegangen.

Die vorliegende Erfindung beschreibt somit ein Verfahren zur parallelen Detektion des Methylierungszustandes genomicscher DNA. Dabei sollen mehrere Cytosin-Methylierungen in einer DNA-Probe gleichzeitig analysiert werden. Dazu werden die folgenden Verfahrensschritte nacheinander ausgeführt:

Zuerst wird eine genomiche DNA Probe derart chemisch behandelt, daß an der 5'-Position unmethylierte Cytosinbasen in Uracil, Thymin oder eine andere vom Hybridisierungsverhalten her dem Cytosin unähnliche Base verwandelt werden. Bevorzugt wird dazu die oben beschriebene Behandlung genomicischer DNA mit Bisulfit (Hydrogensulfit, Disulfit) und anschließender alkalischer Hydrolyse verwendet, die zu einer Umwandlung nicht methylierter Cytosin-Nukleobasen in Uracil führt.

In einem zweiten Verfahrensschritt werden aus der vorbehandelten genomicischen DNA mehr als zehn unterschiedliche Fragmente gleichzeitig durch Verwendung von synthetischen Oligonukleotiden als Primer amplifiziert, wobei mehr als doppelt so viele Fragmente als statistisch zu erwarten aus an der Genregulation beteiligten, transkribierten und/oder translatierten Sequenzen stammen. Dies kann mittels verschiedener Methoden erreicht werden.

In einer bevorzugten Variante des Verfahrens enthält mindestens eines der für die Amplifikation verwendeten Oligonukleotide weniger Nukleobasen als es statistisch für

eine sequenzspezifische Hybridisierung an die chemisch behandelte genomische DNA Probe erforderlich wäre, was zur Amplifikation mehrerer Fragmente gleichzeitig führen kann. Dabei ist die Gesamtzahl der in diesem Oligonukleotid enthaltenen Nukleobasen kleiner als 17. In einer besonders bevorzugten Variante des Verfahrens ist die Anzahl der in diesem Oligonukleotid enthaltenen Nukleobasen kleiner als 14.

In einer weiteren, bevorzugten Variante des Verfahrens werden für die Amplifikation mehr als 4 Oligonukleotide mit unterschiedlicher Sequenz gleichzeitig in einem Reaktionsgefäß verwendet. In einer besonders bevorzugten Varianten werden zur Herstellung eines komplexen Amplifikates mehr als 26 verschiedene Oligonukleotide gleichzeitig verwendet. In einer besonders bevorzugten Variante des Verfahrens stammt mehr als eine doppelt so hohe Anzahl, wie statistisch zu erwarten, aus an der Regulation von Genen beteiligten Genomabschnitten, z.B. Promotoren und Enhancern, stammt, als bei einer rein zufälligen Wahl der Oligonukleotidsequenzen zu erwarten wäre. In einer weiteren besonders bevorzugten Variante des Verfahrens stammt mehr als eine doppelt so hohe Anzahl der amplifizierten Fragmente aus Genomabschnitten, die in mindestens einer Zelle des jeweiligen Organismus in mRNA transkribiert werden, oder aber aus nach der Transkription in mRNA gespliceten Genomabschnitten (Exons), als bei einer rein zufälligen Wahl der Oligonukleotidsequenzen zu erwarten wäre.

In einer weiteren besonders bevorzugten Variante des Verfahrens stammt mehr als eine doppelt so hohe Anzahl der amplifizierten Fragmente aus Genomabschnitten, welche für Teile einer oder mehrerer Genfamilien kodieren, oder aber sie stammen aus Genomabschnitten, welche für sogenannte „matrix attachment sites“ (MARs)-charakteristische Sequenzen enthalten, als bei einer rein zufälligen Wahl der Oligonukleotidsequenzen zu erwarten wäre.

In einer weiteren besonders bevorzugten Variante des Verfahrens stammt mehr als eine doppelt so hohe Anzahl der amplifizierten Fragmente aus Genomabschnitten, welche als sogenannte „boundary elements“ die Verpackungsdichte des Chromatins organisieren, oder aber sie stammen aus multiple drug resistance gene“ (MDR)-

Promotoren oder kodierenden Regionen, als bei einer rein zufälligen Wahl der Oligonukleotidsequenzen zu erwarten wäre.

In einer weiteren, besonders bevorzugten Variante des Verfahrens werden zur Amplifikation der beschriebenen Fragmente zwei Oligonukleotide oder zwei Klassen von Oligonukleotiden verwendet, von denen eines oder eine Klasse außer im Kontext CpG oder CpNpG zwar die Base C enthalten kann, nicht aber die Base G und von denen das andere oder die andere Klasse außer im Kontext CpG oder CpNpG zwar die Base G, nicht aber die Base C enthalten kann.

In einer weiteren bevorzugten Variante des Verfahrens wird die Amplifikation mittels zweier Oligonukleotide durchgeführt, von denen eines eine vier bis sechzehn Basen lange Sequenz enthält, die zu einer solchen DNA komplementär ist oder dieser entspricht, wie sie entstehen würde, wenn ein ebenso langes DNA Fragment, an welches einer der Faktoren

AhR/Arnt	aryl hydrocarbon receptor/aryl hydrocarbon receptor nuclear translocator
Arnt	aryl hydrocarbon receptor nuclear translocator
AML-1a	CBF <sub>A</sub> 2; core-binding factor, runt domain, alpha subunit 2 (acute myeloid leukemia 1; aml1 oncogene)
AP-1	activator protein-1 (AP-1); Synonyme: c-Jun
C/EBP	CCAAT/enhancer binding protein
C/EBP <sub>alpha</sub>	CCAAT/enhancer binding protein (C/EBP), alpha
C/EBP <sub>beta</sub>	CCAAT/enhancer binding protein (C/EBP), beta
CDP	CUTL1; cut (Drosophila)-like 1 (CCAAT displacement protein)
CDP	CUTL1; cut (Drosophila)-like 1 (CCAAT displacement protein)
CDP CR1	complement component (3b/4b) receptor 1
CDP CR3	complement component (3b/4b) receptor 3
CHOP-C/EBP <sub>alpha</sub>	DDIT; DNA-damage-inducible transcript 3/CCAAT/enhancer binding protein (C/EBP), alpha
c-Myc/Max	avian myelocytomatis viral oncogene/MYC-ASSOCIATED FACTOR X
CREB	cAMP responsive element binding protein
CRE-BP1	CYCLIC AMP RESPONSE ELEMENT-BINDING PROTEIN 2, CREB2, CREBP1; now ATF2; activating transcription factor 2
CRE-BP1/c-Jun	activator protein-1 (AP-1); Synonyme: c-Jun

CREB	MP responsive element binding protein
E2F	E2F transcription factor (originally identified as a DNA-binding protein essential E1A-dependent activation of the adenovirus E2 promoter)
E47	transcription factor 3 (E2A immunoglobulin enhancer binding factors E12/E47)
E47	transcription factor 3 (E2A immunoglobulin enhancer binding factors E12/E47)
Egr-1	early growth response 1
Egr-2	early growth response 2 (Krox-20 ( <i>Drosophila</i> ) homolog)
ELK-1	ELK1, member of ETS (environmental tobacco smoke) oncogene family
Freac-2	FKHL6; forkhead ( <i>Drosophila</i> )-like 6; FORKHEAD-RELATED ACTIVATOR 2; FREAC2
Freac-3	FKHL7; forkhead ( <i>Drosophila</i> )-like 7; FORKHEAD-RELATED ACTIVATOR 3; FREAC3
Freac-4	FKHL8; forkhead ( <i>Drosophila</i> )-like 8; FORKHEAD-RELATED ACTIVATOR 4; FREAC4
Freac-7	FKHL11; forkhead ( <i>Drosophila</i> )-like 9; FORKHEAD-RELATED ACTIVATOR 7; FREAC7
GATA-1	GATA-binding protein 1/Enhancer-Binding Protein GATA1
GATA-1	GATA-binding protein 1/Enhancer-Binding Protein GATA1
GATA-1	GATA-binding protein 1/Enhancer-Binding Protein GATA1
GATA-2	GATA-binding protein 2/Enhancer-Binding Protein GATA2
GATA-3	GATA-binding protein 3/Enhancer-Binding Protein GATA3
GATA-X	
HFH-3	FKHL10; forkhead ( <i>Drosophila</i> )-like 10; FORKHEAD-RELATED ACTIVATOR 6; FREAC6
HNF-1	TCF1; transcription factor 1, hepatic; LF-B1, hepatic nuclear factor (HNF1), albumin proximal factor
HNF-4	hepatocyte nuclear factor 4
IRF-1	interferon regulatory factor 1
ISRE	interferon-stimulated response element
Lmo2 complex	LIM domain only 2 (rhombotin-like 1)
MEF-2	MADS box transcription enhancer factor 2, polypeptide A (myocyte enhancer factor 2A)
MEF-2	MADS box transcription enhancer factor 2, polypeptide A (myocyte enhancer factor 2A)
myogenin/NF-1	Myogenin (myogenic factor 4)/Neurofibromin 1; NEUROFIBROMATOSIS, TYPE I
MZF1	ZNF42; zinc finger protein 42 (myeloid-specific retinoic acid-responsive)
MZF1	ZNF42; zinc finger protein 42 (myeloid-specific retinoic acid-responsive)
NF-E2	NFE2; nuclear factor (erythroid-derived 2), 45kD
NF-kappaB (p50)	nuclear factor of kappa light polypeptide gene enhancer in B-cells p50 subunit
NF-kappaB (p65)	nuclear factor of kappa light polypeptide gene enhancer in B-

	cells p65 subunit
NF-kappaB	nuclear factor of kappa light polypeptide gene enhancer in B-cells
NF-kappaB	nuclear factor of kappa light polypeptide gene enhancer in B-cells
NRSF	NEURON RESTRICTIVE SILENCER FACTOR; REST; RE1-silencing transcription factor
Oct-1	OCTAMER-BINDING TRANSCRIPTION FACTOR 1; POU2F1; POU domain, class 2, transcription factor 1
Oct-1	OCTAMER-BINDING TRANSCRIPTION FACTOR 1; POU2F1; POU domain, class 2, transcription factor 1
Oct-1	OCTAMER-BINDING TRANSCRIPTION FACTOR 1; POU2F1; POU domain, class 2, transcription factor 1
Oct-1	OCTAMER-BINDING TRANSCRIPTION FACTOR 1; POU2F1; POU domain, class 2, transcription factor 1
Oct-1	OCTAMER-BINDING TRANSCRIPTION FACTOR 1; POU2F1; POU domain, class 2, transcription factor 1
Oct-1	OCTAMER-BINDING TRANSCRIPTION FACTOR 1; POU2F1; POU domain, class 2, transcription factor 1
P300	E1A (adenovirus E1A oncoprotein)-BINDING PROTEIN, 300-KD
P53	tumor protein p53 (Li-Fraumeni syndrome); TP53
Pax-1	paired box gene 1
Pax-3	paired box gene 3 (Waardenburg syndrome 1)
Pax-6	paired box gene 6 (aniridia, keratitis)
Pbx.1b	pre-B-cell leukemia transcription factor
Pbx-1	pre-B-cell leukemia transcription factor 1
RORalpha2	RAR-RELATED ORPHAN RECEPTOR ALPHA; RETINOIC ACID-BINDING RECEPTOR ALPHA
RREB-1	ras responsive element binding protein 1
SP1	simian-virus-40-protein-1
SP1	simian-virus-40-protein-1
SREBP-1	sterol regulatory element binding transcription factor 1
SRF	serum response factor (c-fos serum response element-binding transcription factor)
SRY	sex determining region Y
STAT3	signal transducer and activator of transcription 1, 91kD
Tal-1alpha/E47	T-cell acute lymphocytic leukemia 1/transcription factor 3 (E2A immunoglobulin enhancer binding factors E12/E47)
TATA	cellular and viral TATA box elements
Tax/CREB	Transiently-expressed axonal glycoprotein/cAMP responsive element binding protein
Tax/CREB	Transiently-expressed axonal glycoprotein/cAMP responsive element binding protein
TCF11/MafG	v-maf musculoaponeurotic fibrosarcoma (avian) oncogene family, protein G
TCF11	Transcription Factor 11; TCF11; NFE2L1; nuclear factor (erythroid-derived 2)-like 1
USF	upstream stimulating factor
Whn	winged-helix nude

X-BP-1  
YY1

X-box binding protein 1 oder  
ubiquitously distributed transcription factor belonging to  
the GLI-Kruppel class of zinc finger proteins

bindet, derart chemisch behandelt würde, daß an der 5'-Position unmethylierte Cytosinbasen in Uracil, Thymidin oder eine andere vom Hybridisierungsverhalten her dem Cytosin unähnliche Base verwandelt werden.

In einer weiteren bevorzugten Variante des Verfahrens wird die Amplifikation mittels zweier Oligonukleotide oder zweier Klassen von Oligonukleotiden durchgeführt, von denen eines oder die eine Klasse die vier bis sechzehn Basen lange Sequenz enthält, welche zu einer solchen DNA komplementär ist oder dieser entspricht, wie sie entstehen würde, wenn ein ebenso langes DNA Fragment, welches über seine Sequenz oder Sekundärstruktur die spezifische Lokalisierung von Genom/Chromatinabschnitten innerhalb des Zellkerns herbeiführen kann, derart chemisch behandelt würde, daß an der 5'-Position unmethylierte Cytosinbasen in Uracil, Thymidin oder eine andere vom Hybridisierungsverhalten her dem Cytosin unähnliche Base verwandelt werden.

In einer weiteren bevorzugten Variante des Verfahrens wird die Amplifikation mittels zweier Oligonukleotide oder zweier Klassen von Oligonukleotiden durchgeführt, von denen eines oder die eine Klasse eine der Sequenzen

TCGCGTGT, TACACGCGA, TGTACGCGA, TCGCGTACA,  
TTGCGTGT, AACACGCAA, GGTACGTAA, TTACGTACC,  
TCGCGTGT, AACACGCGA, GGTACGCGA, TCGCGTACC,  
TTGCGTGT, TACACGCAA, TGTACGTAA, TTACGTACA,  
TACGTG, CACGTA, TACGTG, CACGTA,

ATTGCGTGT, ACACGCAAT, GTACGTAAT, ATTACGTAC,  
ATTGCGTGA, TCACGCAAT, TTACGTAAT, ATTACGTAAT,  
ATCGCGTGA, TCACGCGAT, TTACGCGAT, ATCGCGTAA,  
ATCGCGTGT, ACACGCGAT, GTACGCGAT, ATCGCGTAC,  
TGTGGT, ACCACA, ATTATA, TATAAT,

TGAGTTAG, CTAACTCA, TTGATTAA, TAAATCAA,  
TGATTTAG, CTAAATCA, TTGAGTTA, TAACTCAA,

TTTGGT, ACCAAA, ATTAAA, TTTAAT,  
TGTGGA, TCCACA, TTTATA, TATAAA ,  
TTTGGT, TCCAAA, TTTAAA, TTTAAA,  
TGTGGT, ACCACA, ATTATA, TATAAT,  
  
ATTAT, ATAAT, GTAAT, ATTAC,  
ATTGT, ACAAT, GTAAT, ATTAC,  
  
GAAAG, CTTTC, TTTTT, AAAAAA,  
GTAAT, ATTAC, ATTGT, ACAAT,  
GAAAT, ATTTTC, ATTTTT, AAAAT,  
GTAAG, CTTAC, TTTGT, ACAAA,  
TTAATAATCGAT, ATCGATTATTAA, ATCGATTATTGG, CCAATAATCGAT,  
ATCGATTA, TAATCGAT, TAATCGAT, ATCGATTA,  
  
ATCGATCGG, CCGATCGAT, TCGATCGAT, ATCGATCGA,  
ATCGATCGT, ACGATCGAT, GCGATCGAT, ATCGATCGC,  
  
TATCGATA, TATCGATA, TATCGGTG, CACCGATA,  
TATTAATA, TATTAATA, TATTGGTG, CACCAATA,  
  
GTGTAATATTT, AAATATTACAC, GGGTATTGTAT, ATACAATAACCC,  
GTGTAATTTTT, AAAAATTACAC, GGGGATTGTAT, ATACAATCCCC,  
ATGTAATTTTT, AAAAATTACAT, GGGGATTGTAT, ATACAATCCCC,  
ATGTAATATTT, AAATATTACAT, GGGTATTGTAT, ATACAATAACCC,  
ATTACGTGGT, ACCACGTAAT, ATTACGTGGT, ACCACGTAAT,  
TGACGTA, TTACGTCA, TTACGTTA, TAACGTA,  
TGACGTTA, TAACGTCA, TGACGTTA, TAACGTCA,  
TTACGTTA, TTACGTAA, TTACGTTA, TTACGTTA,  
TGACGTTA, TAACGTCA, TAACGTTA, TAACGTTA,  
  
TGACGT, ACGTCA, GCGTTA, TAACGC,  
TGACGT, ACGTCA, ACGTTA, TAACGT,  
TTTCGCGT, ACGCGAAA, GCGCGAAA, TTTCGCGC,  
TTTGGCGT, ACGCCAAA, GCGTTAAA, TTTAACGC,  
  
TAGGTGTTA, TAACACCTA, TAATATTG, CAAATATTA,  
TAGGTGTTT, AAACACCTA, GAATATTG, CAAATATTC,  
  
GTAGGTGG, CCACCTAC, TTATTGT, ACAAAATAA,  
GTAGGTGT, ACACCTAC, ATATTGT, ACAAAATAT,  
  
TGC GTGGCGG, CCGCCCACGCA, TCGTTTACGTA, TACGTAAACGA,  
TGC GTGGCGT, ACGCCCACGCA, ACGTTTACGTA, TACGTAAACGT,  
  
TGC GTAGGC GT, ACGCCTACGCA, ACGTTTACGTA, TACGTAAACGT,  
TGC GTAGGC GG, CCGCCTACGCA, TCGTTTACGTA, TACGTAAACGA,  
ATAGGAAGT, ACTTCCTAT, ATTTTTGT, ACAAAAAT,

TCGGAAGT, ACTTCCGA, ATTTTCGG, CCGAAAAT,  
TCGGAAGT, ACTTCCGA, GTTTTCGG, CCGAAAAC,  
TCGGAAAT, ATTTCCGA, ATTTTCGG, CCGAAAAT,  
TCGGAAAT, ATTTCCGA, GTTTTCGG, CCGAAAAC,  
GTAAATAA, TTATTTAC, TTGTTTAT, ATAAACAA,  
GTAAATAAATA, TATTTATTTAC, TGTTTATTTAT, ATAAATAAACAA,  
  
AAAGTAAATA, TATTTACTTT, TGTTTATTTT, AAAATAAACAA,  
AATGTAAATA, TATTTACATT, TGTTTATATT, AATATAAACAA,  
TAAGTAAATA, TATTTACTTA, TGTTTATTTA, TAAATAAACAA,  
TATGTAAATA, TATTTACATA, TGTTTATATA, TATATAAACAA,  
  
ATAAATA, TATTTAT, TGTTTAT, ATAAACAA,  
ATAAATA, TATTTAT, TATTTAT, ATAAATA,  
GATA, TATC, TATT, AATA,  
  
TAGATAA, TTATCTA, TTATTTG, CAAATAA,  
TTGATAA, TTATCAA, TTATTAG, CTAATAA,  
GATAA, TTATC, TTATT, AATAA,  
  
GATG, CATC, TATT, AATA,  
  
GATAG, CTATC, TTATT, AATAA,  
GATAAG, CTTATC, TTTATT, AATAAA,  
  
TGTTTATTTA, TAAATAAACAA, TAAATAAATA, TATTTATTTA,  
TGTTTGTAA, TAAACAAACA, TAAATAAATA, TATTTATTTA,  
TATTTATTTA, TAAATAAATA, TAAATAAATA, TATTTATTTA,  
TATTTGTAA, TAAACAAATA, TAAATAAATA, TATTTATTTA,  
  
GTTAATGATT, AATCATTAAAC, AATTATTAAT, ATTAATAATT,  
GTTAATTATT, AATAATTAAC, AATAATTAAT, ATTAATTATT,  
GTTAATTAAT, ATTAATTAAC, ATTAATTAAT, ATTAATTAAT,  
GTTAATGAAT, ATTCAATTAAAC, ATTTATTAAT, ATTAATAAAAT,  
  
TAAAGTTA, TAAACTTA, TGAATTTG, CAAAATTCA,  
TAAAGGTAA, TAACCTTA, TGATTTTG, CAAAAATCA,  
  
AAAGTGAAATT, AATTCACTTT, GGTTTATTTT, AAAATAAAACC,  
AAAGCGAAATT, AATTCGCTTT, GGTTTCGTTT, AAAACGAAACC,  
  
TAGTTTATTTTTT, AAAAAAATAAAACTA, GGGAAAGTGAAATTG,  
CAATTTCACTTCCC,  
TAGTTTATTTTTT, AAAAAAATAAAACTA, GGAAAAGTGAAATTG,  
CAATTTCACTTTCC,  
TAGTTTTTTTTT, AAAAAAAAAAAACTA, GGAAAAGAGAAATTG,  
CAATTTCTCTTTCC,

TAGTTTTTTTTTTT, AAAAAAAAACCA, GGGAAAGAGAAATTG,  
CAATTCCTTCCC,  
TAGGTG, CACCTA, TATTTG, CAAATA,

TTTTAAAAATAATTTT, AAAATTATTTTAAAA, AGGGTTATTTTAGAG,  
CTCTAAAAATAACCCT,  
TTTTAAAAATAATTTT, AAAATTATTTTAAAA, GGAGTTATTTTAGAG,  
CTCTAAAAATAACTCC,  
TTTTAAAAATAATTTT, AAAATTATTTTAAAA, AGAGTTATTTTAGAG,  
CTCTAAAAATAACTCT,  
TTTTAAAAATAATTTT, AAAATTATTTTAAAA, GGGGTTATTTTAGAG,  
CTCTAAAAATAACCCC,

TGTTATTAAGAGAAA, TTTCTATTTTAATAACA,  
TTTTTATTTTAGTAATA, TATTACTAAAAATAAAAA,  
TGTTATTAAGAGAAA, ATTCTATTTTAATAACA,  
GTTTTATTTTAGTAATA, TATTACTAAAAATAAAC,  
TTTGGTAT, ATACCAAA, GTGTTAAA, TTTAACAC  
GGGGA, TCCCC, TTTTT, AAAAA,

TAGGGG, CCCCTA, TTTTTA, TAAAAAA,  
GAGGGG, CCCCTC, TTTTTT, AAAAAAA,  
TGTTGAGTTAT, ATAACCAACA, ATGATTAGTA, TACTAAATCAT,  
TGTTGATTTAT, ATAATCAACA, GTGAGTTAGTA, TACTAAACTCAC,  
TGTTGAGTTAT, ATAACCAACA, ATGATTAGTA, TACTAAATCAT,  
TGTTGATTTAT, ATAATCAACA, GTGAGTTAGTA, TACTAAACTCAC,  
GGGGATTTT, AAAAATCCCC, GGGAAATTTT, AAAAATTCCC,  
GGGGATTTT, AAAAATCCCC, GGGGATTTT, AAAAATCCCC,  
GGGGATTTT, AAAAATCCCC, GGAAATTTT, AAAAATTCC,  
GGGAATTTT, AAAAATTCCC, GGAAATTTT, AAAAATTCC,  
GGGAATTTT, AAAAATTCCC, GGAAATTTT, AAAAATTCC,  
GGGATTTT, AAAAATCCC, GGGAAATTTT, AAAACTTCCC,  
GGGATTTT, AAAAATCCC, TGGAAAGTTT, AAAACTTCCA,  
TTTAGTATTACGGATAGAGGT, ACCTCTATCCGTAATACTAAA,  
GTTTTGTTCGTGGTGTGAA, TTCAACACCAACGAACAAAAAC,  
TTTAGTATTACGGATAGAGTT, AACTCTATCCGTAATACTAAA,  
GGTTTGTTCGTGGTGTGAA, TTCAACACCAACGAACAAACGCC,  
TTTAGTATTACGGATAGCGGT, ACCGCTATCCGTAATACTAAA,  
GTCGTTGTTCGTGGTGTGAA, TTCAACACCAACGAACAAACGAC,

ATATGTAAAT, ATTTACATAT, ATTTGTATAT, ATATACAAAT,  
TTATGTAAAT, ATTTACATAA, ATTTGTATAA, TTATACAAAT,

GAATATTAA, TAAATATTC, TGAATATTT, AAATATTCA,

GAATATGTA, TACATATTTC, TGTATATT, AAATATACA,  
ATAAT, ATTAT, ATTAT, ATAAT,  
GTAAT, ATTAC, ATTAT, ATAAT,  
AATGTAAAAT, ATTTACATT, ATTTGTATT, AATACAAAT,  
ATTTGTATATT, AATATACAAAT, GGTATGTAAT, ATTTACATACC,  
ATTTGTATATT, AATATACAAAT, AATATGTAAT, ATTTACATATT,  
ATTTGTATATT, AATATACAAAT, AGTATGTAAT, ATTTACATACT,  
ATTTGTATATT, AATATACAAAT, GATATGTAAT, ATTTACATATC,  
AGGAGT, ACTCCT, ATTTTT, AAAAAT,  
GGGAGT, ACTCCC, ATTTTT, AAAAAT,  
GGATATGTTGGGTATGTTT, AACATACCCGAACATATCC,  
GGATATGTTGGGTATGTTT, AACATACCCGAACATATCC,  
GGATATGTTGGGTATGTTT, AACATACCCGAACATATCC;  
AGATATGTTGGGTATGTTT, AACATACCCGAACATATCT,  
TCGTTCGTTTAGATAT, ATATCTAAACGAAACGA,  
ATATTAGAGCGGAACGG, CCGTTCCGCTCTAAATAT,  
CGTTACGGTT, AACCGTAACG, AATCGTGACG, CGTCACGATT,  
CGTTACGGTT, AACCGTAACG, GATCGTGACG, CGTCACGATC,  
CGTTACGTTT, AACCGTAACG, AAGCGTGACG, CGTCACGCTT,  
CGTTACGTTT, AACCGTAACG, GAGCGTGACG, CGTCACGCTC,  
TTTACGTATGA, TCATACGTAAA, TTATGCGTGAA, TTCACGCATAA,  
TTTACGTTGA, TCAAACGTAAA, TTAAGCGTGAA, TTCACGCTTAA,  
TTTACGTTTA, TAAAACGTAAA, TGAAGCGTGAA, TTCACGCTTCA,  
TTTACGTATTA, TAATACGTAAA, TGATGCGTGAA, TTCACGCATCA,  
AATTAATTAA, TTAATTAATT, TTGATTGATT, AATCAATCAA,  
TATTAATTAA, TTAATTAATA, TTGATTGATG, CATCAATCAA,  
TAATTAT, ATAATTA, ATGATTG, CAATCAT,  
TAGGTTA, TAACCTA, TGATTTA, TAAATCA,  
TTTAAATATTTT, AAAATATTTAAA, GGGGGTGGTTGGGG,  
CCCCAAACACCCCC,  
TTTAAATTTTT, AAAATAATTTAAA, GGGGTGGTTGGGG,  
CCCCAAACACCCCC,  
TTTAAATTTTTT, AAAAAAATTTAAA, GGGGGGGTTGGGG,  
CCCCAAACACCCCC,  
TTTAAATAATTTT, AAAATTATTTAAA, GGGGTTGGTTGGGG,  
CCCCAAACACCC,  
GAGGCAGGGG, CCCCGCCTC, TTTCGTTT, AAAACGAAA,

GAGGTAGGG, CCCTACCTC, TTTTGT<sub>TTTT</sub>, AAAACAAAAA,  
AAGGC<sub>GGGG</sub>, CCCCGCC<sub>TT</sub>, TTTCG<sub>TTTT</sub>, AAAACGAAA,  
AAGGTAGGG, CCCTACCTT, TTTTGT<sub>TTTT</sub>, AAAACAAAAA,  
  
GGGGGCGGGGT, ACCCGCCCCC, ATTCG<sub>TTTT</sub>, AAAAACGAAAT,  
GGGGGCGGGGT, ACCCGCCCCC, GTTCG<sub>TTTT</sub>, AAAAACGAAAC,  
TATTATTTAT, ATAAAATAATA, GTGGGGTGATA, TATCACCCCCAC,  
GATTATTTAT, ATAAAATAATC, GTGGGGTGATT, AATCACCCCCAC,  
  
ATTACGTGAT, ATCACGTAAT, ATTACGTGAT, ATCACGTAAT,  
ATTACGTGAT, ATCACGTAAT, GTTACGTGAT, ATCACGTAAC,  
  
TTTATATGG, CCATATAAAA, TTATATAAGG, CCTTATATAA,  
TTATATATGG, CCATATATAA, TTATATATGG, CCATATATAA,  
AAATAAT, ATTATTT, GTTGT<sub>TTT</sub>, AAACAAC,  
AAATTAA, TTAATT, TTAGTT, AAACTAA,  
AAATTAT, ATAATT, GTAGTT, AAACTAC,  
AAATAAA, TTTATT, TTTGT<sub>TTT</sub>, AAACAAA,  
  
ATTTTCGGAAATG, CATTCCGAAAAAT, TATTTCGGGAAAT,  
ATTTCCGAAAATA,  
ATTTTCGGAAATG, CATTCCGAAAAAT, TATTTCGGGAAAT,  
ATTTCCGAAAATA,  
ATTTTCGGAAATG, CATTCCCGAAAAT, TATTTTCGGAAAT,  
ATTTCCGAAAATA,  
ATTTTCGGGAAGTG, CACTTCCC<sub>AAAAT</sub>, TATTTTCGGAAAT,  
ATTTCCGAAAATA,  
  
AATAGATGTT, AACATCTATT, AATATTGTT, AACAAATATT,  
AATAGATGGT, ACCATCTATT, ATTATTGTT, AACAAATAAT,  
  
GTATAAATA, TATTTATAC, TATTATAT, ATATAAATA,  
GTATAAATG, CATTTATAC, TATTATAT, ATATAAATA,  
GTATAAAA, TTTTATAC, TTTTATAT, ATATAAAA,  
GTATAAAAG, CTTTATAC, TTTTATAT, ATATAAAA,  
TTATAAATA, TATTTATAA, TATTATAG, CTATAAATA,  
TTATAAATG, CATTTATAA, TATTATAG, CTATAAATA,  
TTATAAAA, TTTTATAA, TTTTATAG, CTATAAAA,  
TTATAAAAG, CTTTATAA, TTTTATAG, CTATAAAA,  
GGGGGTTGACGTA, TACGTCAACCCCC, TGCGTTAATT<sub>TTTT</sub>,  
AAAAATTAACGCA,  
GGGGGTTGACGTA, TACGTCAACCCCC, TACGT<sub>TTAATT</sub>TTT,  
AAAAATTAACGTA,  
  
TGACGTATATT<sub>TTTT</sub>, AAAAATATACGTCA, GGGGATATGCGTTA,  
TAACGCATATCCCC,  
TGACGTATATT<sub>TTTT</sub>, AAAAATATACGTCA, GGGGATATGCGTTA,  
TAACGCATACCCCC,

ATGATTTAGTA, TACTAAATCAT, TGTTGAGTTAT, ATAACTCAACA,  
GTTAT, ATAAC, ATGAT, ATCAT,

TTACGTGA, TCACGTAA, TTACGTGG, CCACGTAA,  
TTACGTGG, CCACGTAA, TTACGTGG, CCACGTAA,  
TTACGTGG, CCACGTAA, TTACGTGA, TCACGTAA,  
TTACGTGA, TCACGTAA, TTACGTGA, TCACGTAA,  
GACGTT, AACGTC, AGCGTT, AACGCT,

TGACGTGT, ACACGTCA, ATACGTTA, TAACGTAT,  
TGACGTGG, CCACGTCA, TTACGTTA, TAACGTAA,  
CGGTTATTTG, CAAAATAACCG, TAAGATGGTCG oder CGACCATCTTA

enthält, welche zu einer solchen DNA komplementär ist oder dieser entspricht, wie sie entstehen würde, wenn ein ebenso langes DNA Fragment, welches über seine Sequenz oder Sekundärstruktur die spezifische Lokalisierung von Genom/Chromatinabschnitten innerhalb des Zellkerns herbeiführen kann, derart chemisch behandelt würde, daß an der 5'-Position unmethylierte Cytosinbasen in Uracil, Thymidin oder eine andere vom Hybridisierungsverhalten her dem Cytosin unähnliche Base verwandelt werden.

In einer besonders bevorzugten Variante des Verfahrens enthalten die zur Amplifikation verwendeten Oligonukleotide außer den oben definierten Konsensussequenzen mehrere Positionen, an denen entweder irgendeine der drei Basen G, A und T oder irgendeine der Basen C, A und T vorhanden sein kann.

In einer besonders bevorzugten Variante des Verfahrens enthalten die zur Amplifikation verwendeten Oligonukleotide außer einer der oben beschriebenen Konsensussequenzen nur maximal zusätzlich so viele weitere Basen, wie es zur gleichzeitigen Amplifikation von mehr als einhundert verschiedenen Fragmenten pro Reaktion aus der chemisch wie oben behandelten DNA erforderlich ist.

In einem dritten Verfahrensschritt wird nun der Sequenzkontext aller oder eines Teils der in den amplifizierten Fragmenten enthaltenen CpG Dinukleotide oder CpNpG Trinukleotide untersucht.

In einer besonders bevorzugten Variante des Verfahrens erfolgt die Analyse durch Hybridisierung der bereits in der Amplifikation mit einem Fluoreszenzmarker versehenen Fragmente an einen Oligonukleotid- Array (DNA Chip). Der Fluoreszenzmarker kann entweder über die verwendeten Primer oder aber durch ein fluoreszenzmarkiertes Nukleotid (z. B. Cy5-dCTP, kommerziell von Amersham- Pharmacia erhältlich) eingeführt werden.

Dabei hybridisieren komplementäre Fragmente an die jeweiligen auf der Chipoberfläche immobilisierten Oligomere, nicht komplementäre Fragmente werden in einem oder mehreren Waschschritten entfernt. Die Fluoreszenz an den jeweiligen Hybridisierungsstellen auf dem Chip erlaubt dann den Rückschluß auf den Sequenzkontext der in den amplifizierten Fragmenten enthaltenen CpG Dinukleotide oder CpNpG Trinukleotide.

In einer weiteren bevorzugten Variante des Verfahrens werden die amplifizierten Fragmente auf einer Oberfläche immobilisiert und anschließend eine Hybridisierung mit einer kombinatorischen Bibliothek von unterscheidbaren Oligonukleotid- oder PNA- Oligomer-Sonden durchgeführt. Wiederum werden nicht komplementäre Sonden durch einen oder mehrere Waschschritte entfernt. Die hybridisierten Sonden werden entweder über ihre Fluoreszenzmarker detektiert oder in einer weiteren besonders bevorzugten Variante des Verfahrens mittels Matrix-assistierter Laser- Desorptions/Ionisations Massenspektrometrie (MALDI-MS) anhand ihrer eindeutigen Masse nachgewiesen. Dabei werden die Sonnenbibliotheken derart synthetisiert, daß die Masse eines jeden Bestandteils eindeutig seiner Sequenz zugeordnet werden kann.

Die Amplifikate können zudem in einer weiteren bevorzugten Variante des Verfahrens hinsichtlich Ihrer durchschnittlichen Größe durch Veränderung der Kettenverlängerungszeiten im Amplifikationsschritt beeinflußt werden. Da hier vorwiegend kleinere Fragmente (ca. 200-500 Basenpaare) untersucht werden, ist eine Verkürzung der Kettenverlängerungsschritte z. B. einer PCR sinnvoll.

In einer weiteren bevorzugten Variante des Verfahrens werden die Amplifikate durch

Gelelektrophorese aufgetrennt, und die Fragmente im gewünschten Größenbereich werden vor Ihrer Analyse ausgeschnitten. In einer weiteren besonders bevorzugten Variante werden die aus dem Gel ausgeschnittenen Amplifikate unter Verwendung des gleichen Satzes an Primern erneut amplifiziert. Dabei können dann nur noch Fragmente der gewünschten Größe entstehen, da Andere als Templat nicht mehr verfügbar sind.

Ein weiterer Gegenstand der vorliegenden Erfindung ist ein Kit, enthaltend mindestens zwei Primerpaare, Reagenzien und Hilfsstoffe für die Amplifikation und/oder Reagenzien und Hilfsmittel für die chemische Behandlung und/oder eine kombinatorische Sondenbibliothek und/oder einen Oligonukleotid-Array (DNA-Chip), soweit sie für die Durchführung des erfindungsgemäßen Verfahrens erforderlich oder dienlich sind.

Die folgenden Beispiele erläutern die Erfindung.

Beispiele:

Beispiel 1:

Primer zur bevorzugten Amplifikation von CG reichen Regionen im Humangenom

Bei den CG reichen Regionen im Humangenom handelt es sich um sogenannte CpG-islands, die eine regulatorischen Funktion besitzen. Wir definieren CpG Islands derart, dass sie mindestens 500 bp umfassen sowie einen GC-Gehalt von >50% aufweisen , ausserdem ist der Quotient CG/GC > 0,6. Unter diesen Bedingungen liegen 16 Mb als CpG Islands vor. Damit liegen etwa 0,5 % der Genomsequenz in diesen CpG islands, wenn man auch noch jeweils eine Region bis 1000 bp downstream zusätzlich betrachtet. Dieser Überlegung liegen Daten aus der Ensembl Database vom 31.10.00, Quelle Sanger Centre, zugrunde. Die dort verfügbare Sequenz umfasste ca. 3,5 GB, und für die Berechnungen wurden die Repeats maskiert.

Statistisch wäre es bei 12meren zu erwarten, dass sie nur 0,005 mal so häufig an eine der CG-reichen Regionen hybridisieren wie an eine andere beliebige Region im

Genom. Es wurden nun Primer gefunden, welche 1,8 mal häufiger an eine CG reiche Region binden. Zudem ergibt sich mit den entsprechend gefundene Reverse Primer nahezu eine Spezifität für diese CpG islands.

In diesem Beispiel sind die Primer AGTAGTAGTAGT (Seq. ID 1) AAAACAAAAACC (Seq. ID 2) und alternativ AGTAGTAGTAGT (Seq. ID 19) und ACAAAAAACTAAA (seq. ID 20). Das erste Primerpaar führt mindestens zu den Amplifikaten Seq. ID 3 bis 18, das zweite Primerpaar zu den Amplifikaten der Seq. ID 21 bis 31.

### Beispiel 2:

Berechnung der Vorhersage der Anzahl von Amplifikaten in Genomischen Regionen.

Gemäß Anspruch 8 im Patent wird gezeigt mehr als doppelt so viele Amplifikate erstellen zu können, als es statistisch zu erwarten wäre nach Formel 1.

$$F = N * P_s(\text{Primers}) \frac{(P_a(\text{Primers}))}{\log(1 - P_a(\text{Primers}))} [(1 - P_a(\text{Primers}))^M - 1] + N * P_a(\text{Primers}) \frac{(P_s(\text{Primers}))}{\log(1 - P_s(\text{Primers}))} [(1 - P_s(\text{Primers}))^M - 1]$$

Formel 1.

F gibt dabei die Anzahl der Vorhergesagten Amplifikate an, die zu erwarten sind, wenn man N Basen als Datenbasis aus dem Genom betrachtet. P ist die jeweilige Wahrscheinlichkeit für die Hybridisierung eines Primeroligonukleotids, getrennt nach Hybridisierung im Sense- und Antisense-Strang. M ist die maximal zulässige Länge der zu erwartenden Amplifikate.

Die Wahrscheinlichkeit P wird bestimmt durch eine Markov Kette erster Ordnung. Dabei wird die Annahme gemacht, dass die DNA eine Zufallsfolge in Abhängigkeit benachbarter Basen ist. Für die Berechnung einer Markovkette sind die Übergangswahrscheinlichkeiten von benachbarten Basen notwendig. Diese wurden empirisch aus 12% des assemblierten humanen Genoms, das vollständig mit Bisulfit behandelt wurde, ermittelt und in Tabelle 1 zusammengefasst. In Tabelle 2 sind die Übergangswahrscheinlichkeiten für den entsprechenden komplementär reversen

Strang angegeben. Diese ergeben sich durch einfaches Vertauschen der Einträge aus der Tabelle 1.

Tabelle 1

Von\nach	A	C	G	T
A	0.0894	0.0033	0.0722	0.1162
C	0.0	0.0	0.0140	0.0
G	0.0603	0.0036	0.0601	0.0959
T	0.1314	0.0071	0.0736	0.2729

mit

$$P_{bDNA}(A)=0.2811$$

$$P_{bDNA}(C)=0.0140$$

$$P_{bDNA}(G)=0.2199$$

$$P_{bDNA}(T)=0.4850$$

und für den dazu revers-komplementären Strang (durch entsprechendes Austauschen der Einträge)  $P_{rbDNA}(von; nach)$

Tabelle 2

Von\nach	A	C	G	T
A	0.2729	0.0959	0.0	0.1162
C	0.0736	0.0601	0.0140	0.0722
G	0.0071	0.0036	0.0	0.0033
T	0.1314	0.0603	0.0	0.0894

$$P_{rbDNA}(A)=0.4850$$

$$P_{rbDNA}(C)=0.2199$$

$$P_{rbDNA}(G)=0.0140$$

$$P_{rbDNA}(T)=0.2811$$

Damit hängt die Wahrscheinlichkeit, dass sich für einen Primer *PrimE* (mit der

Basenfolge  $B_1 B_2 B_3 B_4 \dots$ ; z.B. ATTG...) eine perfekte Basenpaarung ergibt, von der genauen Abfolge der Basen ab und ergibt sich als das Produkt:

$$P_{3s}(PrimE) = P_{rbDNA}(B_1) \frac{P_{rbDNA}(B_1; B_2)}{P_{rbDNA}(B_1)} \frac{P_{rbDNA}(B_2; B_3)}{P_{rbDNA}(B_2)} \frac{P_{rbDNA}(B_3; B_4)}{P_{rbDNA}(B_3)} \dots$$

(Bisulfit-DNA-Strang)

$$P_{3a}(PrimE) = P_{bDNA}(B_1) \frac{P_{bDNA}(B_1; B_2)}{P_{bDNA}(B_1)} \frac{P_{bDNA}(B_2; B_3)}{P_{bDNA}(B_2)} \frac{P_{bDNA}(B_3; B_4)}{P_{bDNA}(B_3)} \dots$$

(anti-sense-Strang zu einem Bisulfit-DNA-Strang);

für einen Primer *Prim* auf dem sense-Strang ergeben sich

$$N * P_s(Prim)$$

perfekte Basenpaarungen - Werden mehrere Primer (*PrimU*, *PrimV*, *PrimW*, *PrimX*, etc.) gleichzeitig verwendet, ergibt sich als Wahrscheinlichkeit für eine perfekte Basenpaarung auf dem sense-Strang an einer gegebenen Position:

$$\begin{aligned} P_s(Primers) &= P_s(PrimU) \\ &+ (1 - P_s(PrimU)) P_s(PrimV) \\ &+ (1 - P_s(PrimU))(1 - P_s(PrimV)) P_s(PrimW) \\ &+ (1 - P_s(PrimU))(1 - P_s(PrimV))(1 - P_s(PrimW)) P_s(PrimX) \\ &+ \dots \end{aligned}$$

(*PrimU*, *PrimV*, *PrimW*... sind hier verschiedene Primer mit unterschiedlichen Basenpaarungen)

und damit als Anzahl der zu erwartenden perfekten Basenpaarungen mit irgendeinem der Primer

$$N * P_s(Primers)$$

Für die Bestimmung von  $P_a(Primers)$  auf dem anti-sense-Strang werden die analogen Gleichungen verwendet.

Für das Beispiel mit zwei Primern (einem sense-Primer und einem antisense-Primer) ergeben sich folgende Wahrscheinlichkeiten:

$$P(AGTAGTAGTAGT) = 0.000000860027$$

$$P(AACAAAAACTAA) = 0.000030005828$$

Auf den CpG-Islands, die insgesamt ca. 30.000.000 Basen enthalten, erwartet man eine Häufigkeit von Hybridisierungen für:

AGTAGTAGTAGT: 25.80 auf dem sense Strang

AACAAAAAACTAA: 900.17 auf dem komplementär reversen Strang.

Auf den jeweils anderen Strängen können die Primer nicht hybridisieren, da auf dem sense-Strang durch die Bisulfitbehandlung keine Cs außerhalb des Kontextes CG auftreten und entsprechend komplementär auf dem antisense-Strang.

Ein Amplifikat entsteht genau dann, wenn bei einer perfekten Basenpaarung auf dem sense-Strang innerhalb der maximalen Fragmentlänge  $M$  ein Primer auf dem Gegenstrang eine perfekte Basenpaarung bildet, die Wahrscheinlichkeit dafür ist

$$P_a(\text{Primers}) \sum_{i=0}^{M-2} (1 - P_a(\text{Primers}))^i ;$$

für große  $M$  und kleine  $P_a(\text{Primers})$  wird dieses durch folgenden Ausdruck berechnet:

$$\frac{P_a(\text{Primers})}{\log(1 - P_a(\text{Primers}))} [(1 - P_a(\text{Primers}))^M - 1] ;$$

für die Gesamtzahl  $F$  der Amplifikate, die durch die Amplifikation beider Stränge zu erwarten sind, ergibt sich damit

$$F = N * P_s(\text{Primers}) \frac{(P_a(\text{Primers}))}{\log(1 - P_a(\text{Primers}))} [(1 - P_a(\text{Primers}))^M - 1] + N * P_a(\text{Primers}) \frac{(P_s(\text{Primers}))}{\log(1 - P_s(\text{Primers}))} [(1 - P_s(\text{Primers}))^M - 1] \quad \text{Formel 1}$$

Für das oben angegebene Beispiel ergeben sich für die CpG-Islands mit 30 Mega Basen 3.0498 Amplifikate. Wir können jedoch zeigen (siehe Beispiel 1), dass man mit Primern, die für bestimmte Regionen spezifisch sind, mehr als statistisch vorhergesagte Amplifikate erzeugen kann.

### Patentansprüche

1. Verfahren zur parallelen Detektion des Methylierungszustandes von genomischer DNA, dadurch gekennzeichnet, daß man folgende Schritte ausführt:
  - a) in einer genomischen DNA Probe wandelt man durch chemische Behandlung an der 5'-Position unmethylierte Cytosinbasen in Uracil, Thymidin oder eine andere vom Hybridisierungsverhalten her dem Cytosin unähnliche Base um;
  - b) aus dieser chemisch behandelten genomischen DNA amplifiziert man mehr als zehn unterschiedliche Fragmente, die jeweils weniger als 2000 Basenpaare lang sind, gleichzeitig durch Verwendung von synthetischen Oligonukleotiden als Primer, wobei diese Primer jeweils Sequenzen aus an der Genregulation beteiligten und/oder transkribierten und/oder translatierten genomischen Sequenzen enthalten, wie sie nach einer Behandlung gemäß Schritt a) vorliegen würden;
  - c) man bestimmt den Sequenzkontext aller oder eines Teils der in den amplifizierten Fragmenten enthaltenen CpG Dinukleotide oder CpNpG Trinukleotide.
2. Verfahren nach Anspruch 1, dadurch gekennzeichnet, daß man die chemische Behandlung mittels einer Lösung eines Bisulfits, Hydrogensulfits oder Disulfits durchführt.
3. Verfahren nach Anspruch 1 oder 2, dadurch gekennzeichnet, daß mindestens eines der in Schritt b) verwendeten Oligonukleotide weniger Nukleobasen enthält als es statistisch für eine sequenzspezifische Hybridisierung an die chemisch behandelte genomische DNA Probe erforderlich wäre.
4. Verfahren nach einem der Ansprüche 1 bis 3, dadurch gekennzeichnet, daß mindestens eines der in Schritt b) des Anspruchs 1 verwendeten Oligonukleotide

kürzer als 18 Nukleobasen ist.

5. Verfahren nach einem der Ansprüche 1 bis 3, dadurch gekennzeichnet, daß mindestens eines der in Schritt b) des Anspruchs 1 verwendeten Oligonukleotide kürzer als 15 Nukleobasen ist.
6. Verfahren nach Anspruch 1 oder 2, dadurch gekennzeichnet, daß man in Schritt b) des Anspruchs 1 mehr als 4 verschiedene Oligonukleotide gleichzeitig für die Amplifikation verwendet.
7. Verfahren nach Anspruch 1 oder 2, dadurch gekennzeichnet, daß man in Schritt b) des Anspruchs 1 mehr als 26 verschiedene Oligonukleotide gleichzeitig für die Amplifikation verwendet.
8. Verfahren nach einem der vorangehenden Ansprüche, dadurch gekennzeichnet, daß in Schritt b) des Anspruchs 1 mehr als doppelt so viele amplifizierte Fragmente als nach berechnet nach Formel 1 aus an der Regulation von Genen beteiligten Genomabschnitten, wie Promotoren und Enhancern, stammt, als bei einer rein zufälligen Wahl der Oligonukleotidsequenzen zu erwarten wäre, oder aber deren Anteil an den insgesamt nachweisbaren Fragmenten mehr als doppelt so hoch ist wie berechnet nach Formel 1,

$$F = N * P_s(\text{Primers}) \frac{(P_a(\text{Primers}))}{\log(1 - P_a(\text{Primers}))} [(1 - P_a(\text{Primers}))^M - 1] + N * P_a(\text{Primers}) \frac{(P_s(\text{Primers}))}{\log(1 - P_s(\text{Primers}))} [(1 - P_s(\text{Primers}))^M - 1]$$

Formel 1

wobei man die Berechnung wie folgt durchführt:

bei der mit Bisulfit behandelten DNA kann C nur noch im Kontext CG auftreten, so wird angenommen, daß die primäre DNA eine Zufallsfolge mit Abhängigkeit direkt benachbarter Basen ist (Markov-Kette erster Ordnung); die empirisch aus der Datenbank (vollständig methyliert; mit Bisulfit behandelt) ermittelten

paarweisen Basenwahrscheinlichkeiten ergeben sich gleich für beide DNA-Stränge als  $P_{bDNA}(\text{von};\text{nach})$  aus der folgenden Tabelle:

<b>Von\nach</b>	<b>A</b>	<b>C</b>	<b>G</b>	<b>T</b>
A	0.0894	0.0033	0.0722	0.1162
C	0.0	0.0	0.0140	0.0
G	0.0603	0.0036	0.0601	0.0959
T	0.1314	0.0071	0.0736	0.2729

mit

$$P_{bDNA}(A)=0.2811$$

$$P_{bDNA}(C)=0.0140$$

$$P_{bDNA}(G)=0.2199$$

$$P_{bDNA}(T)=0.4850$$

und für den dazu revers-komplementären Strang (durch entsprechendes Austauschen der Einträge)  $P_{rbDNA}(\text{von};\text{nach})$

<b>Von\nach</b>	<b>A</b>	<b>C</b>	<b>G</b>	<b>T</b>
A	0.2729	0.0959	0.0	0.1162
C	0.0736	0.0601	0.0140	0.0722
G	0.0071	0.0036	0.0	0.0033
T	0.1314	0.0603	0.0	0.0894

$$P_{rbDNA}(A)=0.4850$$

$$P_{rbDNA}(C)=0.2199$$

$$P_{rbDNA}(G)=0.0140$$

$$P_{rbDNA}(T)=0.2811$$

damit hängt die Wahrscheinlichkeit, daß sich für einen Primer *PrimE* (mit der Basenfolge B<sub>1</sub> B<sub>2</sub> B<sub>3</sub> B<sub>4</sub> ...; z.B. ATTG...) eine perfekte Basenpaarung ergibt, von der genauen Abfolge der Basen ab und ergibt sich als das Produkt:

$$P_{3s}(PrimE) = P_{rbDNA}(B_1) \frac{P_{rbDNA}(B_1; B_2)}{P_{rbDNA}(B_1)} \frac{P_{rbDNA}(B_2; B_3)}{P_{rbDNA}(B_2)} \frac{P_{rbDNA}(B_3; B_4)}{P_{rbDNA}(B_3)} \dots$$

(Bisulfit-DNA-Strang)

$$P_{3a}(PrimE) = P_{bDNA}(B_1) \frac{P_{bDNA}(B_1; B_2)}{P_{bDNA}(B_1)} \frac{P_{bDNA}(B_2; B_3)}{P_{bDNA}(B_2)} \frac{P_{bDNA}(B_3; B_4)}{P_{bDNA}(B_3)} \dots$$

(anti-sense-Strang zu einem Bisulfit-DNA-Strang);

für einen Primer *Prim* auf dem sense-Strang ergeben sich

$$N * P_s(Prim) ;$$

perfekte Basenpaarungen - Werden mehrere Primer (*PrimU*, *PrimV*, *PrimW*, *PrimX*, etc.) gleichzeitig verwendet, ergibt sich als Wahrscheinlichkeit für eine perfekte Basenpaarung auf dem sense-Strang an einer gegebenen Position:

$$\begin{aligned} P_s(Primers) &= P_s(PrimU) \\ &+ (1 - P_s(PrimU)) P_s(PrimV) \\ &+ (1 - P_s(PrimU))(1 - P_s(PrimV)) P_s(PrimW) \\ &+ (1 - P_s(PrimU))(1 - P_s(PrimV))(1 - P_s(PrimW)) P_s(PrimX) \\ &+ \dots \end{aligned}$$

und damit als Anzahl der zu erwartenden perfekten Basenparungen mit irgendeinem der Primer

$$N * P_s(Primers) ;$$

für die Bestimmung von  $P_a(Primers)$  auf dem anti-sense-Strang werden die analogen Gleichungen verwendet; ein Amplifikat entsteht genau dann, wenn bei einer perfekten Basenpaarung auf dem sense-Strang innerhalb der maximalen Fragmentlänge  $M$  ein Primer auf dem Gegenstrang eine perfekte Basenpaarung bildet, die Wahrscheinlichkeit dafür ist

$$P_a(Primers) \sum_{i=0}^{M-2} (1 - P_a(Primers))^i ;$$

für große  $M$  und kleine  $P_a(Primers)$  wird dieses durch folgenden Ausdruck berechnet:

$$\frac{P_a(Primers)}{\log(1 - P_a(Primers))} [(1 - P_a(Primers))^M - 1] ;$$

für die Gesamtzahl  $F$  der Amplifikate, die durch die Amplifikation beider Stränge zu erwarten sind, ergibt sich damit

$$F = N * P_s(\text{Primers}) \frac{(P_a(\text{Primers}))}{\log(1 - P_a(\text{Primers}))} [(1 - P_a(\text{Primers}))^M - 1] + N * P_a(\text{Primers}) \frac{(P_s(\text{Primers}))}{\log(1 - P_s(\text{Primers}))} [(1 - P_s(\text{Primers}))^M - 1]$$

Formel 1

9. Verfahren nach einem der Ansprüche 1 bis 7, dadurch gekennzeichnet, daß in Schritt b) des Anspruchs 1 mehr als doppelt so viele amplifizierte Fragmente als berechnet entsprechend Anspruch 8 aus Genomabschnitten stammt, die in mindestens einer Zelle des jeweiligen Organismus in mRNA transkribiert werden, als bei einer rein zufälligen Wahl der Oligonukleotidsequenzen zu erwarten wäre, oder aber deren Anteil an den insgesamt nachweisbaren Fragmenten mehr als doppelt so hoch ist als berechnet entsprechend Anspruch 8.
10. Verfahren nach einem der Ansprüche 1 bis 7, dadurch gekennzeichnet, daß in Schritt b) des Anspruchs 1 mehr als doppelt so viele amplifizierte Fragmente als berechnet entsprechend Anspruch 8 aus nach der Transkription in mRNA gesplittenen Genomabschnitten (Exons) stammt, als bei einer rein zufälligen Wahl der Oligonukleotidsequenzen zu erwarten wäre, oder aber deren Anteil an den insgesamt nachweisbaren Fragmenten mehr als doppelt so hoch ist als berechnet entsprechend Anspruch 8.
11. Verfahren nach einem der Ansprüche 1 bis 7, dadurch gekennzeichnet, daß in Schritt b) des Anspruchs 1 mehr als doppelt so viele amplifizierte Fragmente als berechnet entsprechend Anspruch 8 aus Genomabschnitten stammen, welche für Teile einer oder mehrerer Genfamilien kodieren, als bei einer rein zufälligen Wahl der Oligonukleotidsequenzen zu erwarten wäre, oder aber deren Anteil an den insgesamt nachweisbaren Fragmenten mehr als doppelt so hoch ist als berechnet entsprechend Anspruch 8.
12. Verfahren nach einem der Ansprüche 1 bis 7, dadurch gekennzeichnet, daß in Schritt b) des Anspruchs 1 mehr als doppelt so viele amplifizierte Fragmente als berechnet entsprechend Anspruch 8. aus Genomabschnitten stammen, welche für sogenannte „matrix attachment sites“ (MARs)- charakteristische Sequenzen

enthalten, als bei einer rein zufälligen Wahl der Oligonukleotidsequenzen zu erwarten wäre, oder aber deren Anteil an den insgesamt nachweisbaren Fragmenten mehr als doppelt so hoch ist als berechnet entsprechend Anspruch 8.

13. Verfahren nach einem der Ansprüche 1 bis 7, dadurch gekennzeichnet, daß in Schritt b) des Anspruchs 1 mehr als doppelt so viele amplifizierte Fragmente als berechnet entsprechend Anspruch 8 aus Genomabschnitten stammen, welche als sogenannte „boundary elements“ die Verpackungsdichte des Chromatins organisieren, als bei einer rein zufälligen Wahl der Oligonukleotidsequenzen zu erwarten wäre, oder aber deren Anteil an den insgesamt nachweisbaren Fragmenten mehr als doppelt so hoch ist als berechnet entsprechend Anspruch 8.
14. Verfahren nach einem der Ansprüche 1 bis 7, dadurch gekennzeichnet, daß in Schritt b) des Anspruchs 1 mehr als doppelt so viele amplifizierte Fragmente als berechnet entsprechend Anspruch 8 aus „multiple drug resistance gene“ (MDR)-Promotoren oder kodierenden Regionen stammen, als bei einer rein zufälligen Wahl der Oligonukleotidsequenzen zu erwarten wäre oder aber deren Anteil an den insgesamt nachweisbaren Fragmenten mehr als doppelt so hoch ist als berechnet entsprechend Anspruch 8.
15. Verfahren nach einem der vorangehenden Ansprüche, dadurch gekennzeichnet, daß zur Amplifikation der in Anspruch 1 beschriebenen Fragmente zwei Oligonukleotide oder zwei Klassen von Oligonukleotiden verwendet werden, von denen eines oder eine Klasse außer im Kontext CpG oder CpNpG zwar die Base C enthalten kann, nicht aber die Base G und von denen das andere oder die andere Klasse außer im Kontext CpG oder CpNpG zwar die Base G, nicht aber die Base C enthalten kann.
16. Verfahren nach einem der Ansprüche 1 bis 4, dadurch gekennzeichnet, daß die in Anspruch 1 beschriebene Amplifikation mittels zweier Oligonukleotide durchgeführt wird, von denen eines eine vier bis sechzehn Basen lange Sequenz enthält, welche zu einer solchen DNA komplementär ist oder dieser entspricht, wie sie entstehen

würde, wenn ein ebenso langes DNA Fragment, an welches einer der Transkriptionsfaktoren

AhR/Arnt	aryl hydrocarbon receptor/aryl hydrocarbon receptor nuclear translocator
Arnt	aryl hydrocarbon receptor nuclear translocator
AML-1a	CBFA2; core-binding factor, runt domain, alpha subunit 2 (acute myeloid leukemia 1; aml1 oncogene)
AP-1	activator protein-1 (AP-1); Synonyme: c-Jun
C/EBP	CCAAT/enhancer binding protein
C/EBPalpha	CCAAT/enhancer binding protein (C/EBP), alpha
C/EBPbeta	CCAAT/enhancer binding protein (C/EBP), beta
CDP	CUTL1; cut ( <i>Drosophila</i> )-like 1 (CCAAT displacement protein)
CDP	CUTL1; cut ( <i>Drosophila</i> )-like 1 (CCAAT displacement protein)
CDP CR1	complement component (3b/4b) receptor 1
CDP CR3	complement component (3b/4b) receptor 3
CHOP-C/EBPalpha	DDIT; DNA-damage-inducible transcript 3/CCAAT/enhancer binding protein (C/EBP), alpha
c-Myc/Max	avian myelocytomatis viral oncogene/MYC-ASSOCIATED FACTOR X
CREB	cAMP responsive element binding protein
CRE-BP1	CYCLIC AMP RESPONSE ELEMENT-BINDING PROTEIN 2, CREB2, CREBP1; now ATF2; activating transcription factor 2
CRE-BP1/c-Jun	activator protein-1 (AP-1); Synonyme: c-Jun
CREB	MP responsive element binding protein
E2F	E2F transcription factor (originally identified as a DNA-binding protein essential E1A-dependent activation of the adenovirus E2 promoter)
E47	transcription factor 3 (E2A immunoglobulin enhancer binding factors E12/E47)
E47	transcription factor 3 (E2A immunoglobulin enhancer binding factors E12/E47)
Egr-1	early growth response 1
Egr-2	early growth response 2 (Krox-20 ( <i>Drosophila</i> ) homolog)
ELK-1	ELK1, member of ETS (environmental tobacco smoke) oncogene family
Freac-2	FKHL6; forkhead ( <i>Drosophila</i> )-like 6; FORKHEAD-RELATED ACTIVATOR 2; FREAC2
Freac-3	FKHL7; forkhead ( <i>Drosophila</i> )-like 7; FORKHEAD-RELATED ACTIVATOR 3; FREAC3
Freac-4	FKHL8; forkhead ( <i>Drosophila</i> )-like 8; FORKHEAD-RELATED ACTIVATOR 4; FREAC4
Freac-7	FKHL11; forkhead ( <i>Drosophila</i> )-like 9; FORKHEAD-RELATED ACTIVATOR 7; FREAC7

GATA-1	GATA-binding protein 1/Enhancer-Binding Protein GATA1
GATA-1	GATA-binding protein 1/Enhancer-Binding Protein GATA1
GATA-1	GATA-binding protein 1/Enhancer-Binding Protein GATA1
GATA-2	GATA-binding protein 1/Enhancer-Binding Protein GATA1
GATA-3	GATA-binding protein 2/Enhancer-Binding Protein GATA2
GATA-X	GATA-binding protein 3/Enhancer-Binding Protein GATA3
HFH-3	FKHL10; forkhead ( <i>Drosophila</i> )-like 10; FORKHEAD-RELATED ACTIVATOR 6; FREAC6
HNF-1	TCF1; transcription factor 1, hepatic; LF-B1, hepatic nuclear factor (HNF1), albumin proximal factor
HNF-4	hepatocyte nuclear factor 4
IRF-1	interferon regulatory factor 1
ISRE	interferon-stimulated response element
Lmo2 complex	LIM domain only 2 (rhombotin-like 1)
MEF-2	MADS box transcription enhancer factor 2, polypeptide A (myocyte enhancer factor 2A)
MEF-2	MADS box transcription enhancer factor 2, polypeptide A (myocyte enhancer factor 2A)
myogenin/NF-1	Myogenin (myogenic factor 4)/Neurofibromin 1; NEUROFIBROMATOSIS, TYPE I
MZF1	ZNF42; zinc finger protein 42 (myeloid-specific retinoic acid-responsive)
MZF1	ZNF42; zinc finger protein 42 (myeloid-specific retinoic acid-responsive)
NF-E2	NFE2; nuclear factor (erythroid-derived 2), 45kD
NF-kappaB (p50)	nuclear factor of kappa light polypeptide gene enhancer in B-cells p50 subunit
NF-kappaB (p65)	nuclear factor of kappa light polypeptide gene enhancer in B-cells p65 subunit
NF-kappaB	nuclear factor of kappa light polypeptide gene enhancer in B-cells
NF-kappaB	nuclear factor of kappa light polypeptide gene enhancer in B-cells
NRSF	NEURON RESTRICTIVE SILENCER FACTOR; REST; RE1-silencing transcription factor
Oct-1	OCTAMER-BINDING TRANSCRIPTION FACTOR 1; POU2F1; POU domain, class 2, transcription factor 1
Oct-1	OCTAMER-BINDING TRANSCRIPTION FACTOR 1; POU2F1; POU domain, class 2, transcription factor 1
Oct-1	OCTAMER-BINDING TRANSCRIPTION FACTOR 1; POU2F1; POU domain, class 2, transcription factor 1
Oct-1	OCTAMER-BINDING TRANSCRIPTION FACTOR 1; POU2F1; POU domain, class 2, transcription factor 1
Oct-1	OCTAMER-BINDING TRANSCRIPTION FACTOR 1; POU2F1; POU domain, class 2, transcription factor 1
P300	E1A (adenovirus E1A oncoprotein)-BINDING PROTEIN, 300-KD
P53	tumor protein p53 (Li-Fraumeni syndrome); TP53

Pax-1	paired box gene 1
Pax-3	paired box gene 3 (Waardenburg syndrome 1)
Pax-6	paired box gene 6 (aniridia, keratitis)
Pbx 1b	pre-B-cell leukemia transcription factor
Pbx-1	pre-B-cell leukemia transcription factor 1
RORalpha2	RAR-RELATED ORPHAN RECEPTOR ALPHA; RETINOIC ACID-BINDING RECEPTOR ALPHA
RREB-1	ras responsive element binding protein 1
SP1	simian-virus-40-protein-1
SP1	simian-virus-40-protein-1
SREBP-1	sterol regulatory element binding transcription factor 1
SRF	serum response factor (c-fos serum response element-binding transcription factor)
SRY	sex determining region Y
STAT3	signal transducer and activator of transcription 1, 91kD
Tal-1alpha/E47	T-cell acute lymphocytic leukemia 1/transcription factor 3 (E2A immunoglobulin enhancer binding factors E12/E47)
TATA	cellular and viral TATA box elements
Tax/CREB	Transiently-expressed axonal glycoprotein/cAMP responsive element binding protein
Tax/CREB	Transiently-expressed axonal glycoprotein/cAMP responsive element binding protein
TCF11/MafG	v-maf musculoaponeurotic fibrosarcoma (avian) oncogene family, protein G
TCF11	Transcription Factor 11; TCF11; NFE2L1; nuclear factor (erythroid-derived 2)-like 1
USF	upstream stimulating factor
Whn	winged-helix nude
X-BP-1	X-box binding protein 1 oder
YY1	ubiquitously distributed transcription factor belonging to the GLI-Kruppel class of zinc finger proteins

bindet, einer chemischen Behandlung gemäß Anspruch 1 unterzogen würde.

17. Verfahren nach einem der Ansprüche 1 bis 4, dadurch gekennzeichnet, daß man die in Anspruch 1 beschriebene Amplifikation mittels zweier Oligonukleotide durchführt, von denen eine die eine vier bis sechzehn Basen lange Sequenz enthält, welche zu einer solchen DNA komplementär ist oder dieser entspricht, wie sie entstehen würde, wenn ein ebenso langes DNA Fragment, welches über seine Sequenz oder Sekundärstruktur die spezifische Lokalisierung von Genom/Chromatinabschnitten innerhalb des Zellkerns herbeiführen kann, einer chemischen Behandlung gemäß Anspruch 1 unterzogen würde.

18. Verfahren nach einem der Ansprüche 1 bis 4, dadurch gekennzeichnet, daß die in Anspruch 1 beschriebene Amplifikation mittels zweier Oligonukleotide durchgeführt wird, von denen mindestens eine der Sequenzen (von 5' nach 3')

TCGCGTGT, TACACGCGA, TGTACGCGA, TCGCGTACA,  
TTGCGTGT, AACACGCAA, GGTACGTAA, TTACGTACC,  
TCGCGTGT, AACACGCGA, GGTACGCGA, TCGCGTACC,  
TTGCGTGT, TACACGCAA, TGTACGTAA, TTACGTACA,  
TACGTG, CACGTA, TACGTG, CACGTA,

ATTGCGTGT, ACACGCAAT, GTACGTAAT, ATTACGTAC,  
ATTGCGTGA, TCACGCAAT, TTACGTAAT, ATTACGTAA,  
ATCGCGTGA, TCACGCGAT, TTACGCGAT, ATCGCGTAA,  
ATCGCGTGT, ACACGCGAT, GTACGCGAT, ATCGCGTAC,  
TGTGGT, ACCACA, ATTATA, TATAAT,

TGAGTTAG, CTAACTCA, TTGATTTA, TAAATCAA,  
TGATTTAG, CTAAATCA, TTGAGTTA, TAACTCAA,

TTTGGT, ACCAAA, ATTAAA, TTTAAT,  
TGTGGA, TCCACA, TTTATA, TATAAA,  
TTTGGT, TCCAAA, TTTAAA, TTTAAA,  
TGTGGT, ACCACA, ATTATA, TATAAT,

ATTAT, ATAAT, GTAAT, ATTAC,  
ATTGT, ACAAT, GTAAT, ATTAC,

GAAAG, CTTTC, TTTTT, AAAAA,  
GTAAT, ATTAC, ATTGT, ACAAT,  
GAAAT, ATTC, ATTTT, AAAAT,  
GTAAG, CTTAC, TTTGT, AAAAA,  
TTAATAATCGAT, ATCGATTATTAA, ATCGATTATTGG, CCAATAATCGAT,  
ATCGATTA, TAATCGAT, TAATCGAT, ATCGATTA,

ATCGATCGG, CCGATCGAT, TCGATCGAT, ATCGATCGA,  
ATCGATCGT, ACGATCGAT, GCGATCGAT, ATCGATCGC,

TATCGATA, TATCGATA, TATCGGTG, CACCGATA,  
TATTAATA, TATTAATA, TATTGGTG, CACCAATA,

GTGTAATATTT, AAATATTACAC, GGGTATTGTAT, ATACAATACCC,  
GTGTAATTTTT, AAAAATTACAC, GGGGATTGTAT, ATACAATCCCC,  
ATGTAATTTTT, AAAAATTACAT, GGGGATTGTAT, ATACAATCCCC,  
ATGTAATATTT, AAATATTACAT, GGGTATTGTAT, ATACAATACCC,  
ATTACGTGGT, ACCACGTAAT, ATTACGTGGT, ACCACGTAAT,

TGACGTAA, TTACGTCA, TTACGTTA, TAACGTAA,  
TGACGTTA, TAACGTCA, TGACGTTA, TAACGTCA,  
TTACGTAA, TTACGTAA, TTACGTAA, TTACGTAA,  
TGACGTTA, TAACGTCA, TAACGTTA, TAACGTTA,  
  
TGACGT, ACGTCA, GCGTTA, TAACGC,  
TGACGT, ACGTCA, ACGTTA, TAACGT,  
TTTCGCGT, ACGCGAAA, GCGCGAAA, TTTCGCGC,  
TTTGGCGT, ACGCCAAA, GCGTTAAA, TTTAACGC,  
  
TAGGTGTTA, TAACACCTA, TAATATTG, CAAATATTA,  
TAGGTGTTT, AAACACCTA, GAATATTG, CAAATATTC,  
  
GTAGGTGG, CCACCTAC, TTATTTGT, ACAAAATAA,  
GTAGGTGT, ACACCTAC, ATATTTGT, ACAAAATAT,  
  
TGC GTGGCGG, CCGCCCACGCA, TCGTTTACGTA, TACGTAAACGA,  
TGC GTGGCGT, ACGCCCACGCA, ACGTTTACGTA, TACGTAAACGT,  
  
TGC GTAGGC GT, ACGCCTACGCA, ACGTTTACGTA, TACGTAAACGT,  
TGC GTAGGC GG, CCGCCTACGCA, TCGTTTACGTA, TACGTAAACGA,  
ATAGGAAGT, ACTTCCTAT, ATTTTTGT, ACAAAAAT,  
  
TCGGAAGT, ACTTCCGA, ATTTTCGG, CCGAAAAT,  
TCGGAAGT, ACTTCCGA, GTTTTCGG, CCGAAAAC,  
TCGGAAAT, ATTTCCGA, ATTTTCGG, CCGAAAAT,  
TCGGAAAT, ATTTCCGA, GTTTTCGG, CCGAAAAC,  
GTAAATAA, TTATTTAC, TTGTTTAT, ATAACAA,  
GTAAATAAATA, TATTTATTTAC, TGTTTATTTAT, ATAATAAAACA,  
  
AAAGTAAATA, TATTTACTTT, TGTTTATTTT, AAAATAAAACA,  
AATGTAAATA, TATTTACATT, TGTTTATATT, AATATAAAACA,  
TAAGTAAATA, TATTTACTTA, TGTTTATTTA, TAAATAAAACA,  
TATGTAAATA, TATTTACATA, TGTTTATATA, TATATAAAACA,  
  
ATAAATA, TATTTAT, TGTTTAT, ATAACAA,  
ATAAATA, TATTTAT, TATTTAT, ATAATAA,  
GATA, TATC, TATT, AATA,  
  
TAGATAA, TTATCTA, TTATTTG, CAAATAA,  
TTGATAA, TTATCAA, TTATTAG, CTAATAA,  
GATAA, TTATC, TTATT, AATAA,  
  
GATG, CATC, TATT, AATA,  
  
GATAG, CTATC, TTATT, AATAA,  
GATAAG, CTTATC, TTTATT, AATAAA,

TGTTTATTAA, TAAATAAACA, TAAATAAATA, TATTTATTAA,  
TGTTTGTAA, TAAACAAACA, TAAATAAATA, TATTTATTAA,  
TATTTATTAA, TAAATAAATA, TAAATAAATA, TATTTATTAA,  
TATTTGTAA, TAAACAAACA, TAAATAAATA, TATTTATTAA,  
  
GTTAATGATT, AATCATTAAAC, AATTATTAAAT, ATTAATAATT,  
GTTAATTATT, AATAATTAAAC, AATAATTAAAT, ATTAATTATT,  
GTTAATTAAAT, ATTAATTAAAC, ATTAATTAAAT, ATTAATTAAAT,  
GTTAATGAAT, ATTCATTAAAC, ATTTATTAAAT, ATTAATAAAAT,  
  
TAAAGTTA, TAAACTTA, TGAATTGG, CAAAATTCA,  
TAAAGGTAA, TAACCTTA, TGATTTGG, CAAAAATCA,  
  
AAAGTGAATT, AATTCACCTT, GGTTTATTAA, AAAATAAAACC,  
AAAGCGAAATT, AATTCGCTT, GGTTTCGTTT, AAAACGAAACC,  
  
TAGTTTATTTTTT, AAAAAAATAAAACTA, GGGAAAGTGAATTG,  
CAATTTCACCTTCCC,  
TAGTTTATTTTTT, AAAAAAATAAAACTA, GGAAAAGTGAATTG,  
CAATTTCACCTTCCC,  
TAGTTTTTTTTTT, AAAAAAAAAAAACTA, GGAAAAGAGAAATTG,  
CAATTCTCTTTCCC,  
TAGTTTTTTTTTT, AAAAAAAAAAAACTA, GGGAAAGAGAAATTG,  
CAATTCTCTTTCCC,  
TAGGTG, CACCTA, TATTG, CAAATA,  
  
TTTAAAAATAATTAA, AAAATTATTAA, AGGGTTATTAA, TAGAG,  
CTCTAAAATAACCCCT,  
TTTAAAAATAATTAA, AAAATTATTAA, GGAGTTATTAA, TAGAG,  
CTCTAAAATAACTCC,  
TTTAAAAATAATTAA, AAAATTATTAA, AGAGTTATTAA, TAGAG,  
CTCTAAAATAACTCT,  
TTTAAAAATAATTAA, AAAATTATTAA, GGGGTTATTAA, TAGAG,  
CTCTAAAATAACCCCC,  
  
TGTTATTAAAAATAGAAA, TTTCTATTAAATAACA,  
TTTTTATTAAAGTAATA, TATTACTAAAAATAAA,  
TGTTATTAAAAATAGAAT, ATTCTATTAAATAACA,  
GTTTATTAAAGTAATA, TATTACTAAAAATAAAAC,  
TTGGTAT, ATACCAAA, GTGTTAAA, TTAAACAC  
GGGA, TCCCC, TTTT, AAAAA,  
  
TAGGGG, CCCCTA, TTTTTA, TAAAAA,  
GAGGGG, CCCCTC, TTTTTT, AAAAAA,  
TGTTGAGTTAT, ATAACCAACA, ATGATTAGTA, TACTAAATCAT,  
TGTTGATTAT, ATAAATCAACA, GTGAGTTAGTA, TACTAAACTCAC,  
TGTTGAGTTAT, ATAACCAACA, ATGATTAGTA, TACTAAATCAT,  
TGTTGATTAT, ATAAATCAACA, GTGAGTTAGTA, TACTAAACTCAC,

GGGGATTTT, AAAAATCCC, GGGATTTT, AAAAATTCC,  
GGGGATTTT, AAAAATCCC, GGGATTTT, AAAAATCCC,  
GGGGATTTT, AAAAATCCC, GGAAATTTT, AAAAATTCC,  
GGGAATTTT, AAAAATTCC, GGAAATTTT, AAAAATTCC,  
GGGAATTTT, AAAAATTCC, GGAAATTTT, AAAAATTCC,  
GGGATTTTT, AAAAATCCC, GGAAAGTTT, AAAACTTCC,  
GGGAATTTT, AAAAATTCC, GGAAATTTT, AAAAATTCC,  
GGGATTTTT, AAAAATCCC, GGGAAGTTT, AAAACTTCC,  
GGGATTTTTA, TAAAAAATCCC, TGAAAGTTT, AAAACTTCCA,  
TTAGTATTACGGATAGAGGT, ACCTCTATCCGTAACTAAA,  
GTTTTGTTCGTGGTGTGAA, TTCAACACCACGAACAAAAAC,  
TTAGTATTACGGATAGAGTT, AACTCTATCCGTAACTAAA,  
GGTTTGTTCGTGGTGTGAA, TTCAACACCACGAACAAAAAC,  
TTAGTATTACGGATAGCGTT, AACGCTATCCGTAACTAAA,  
GGCGTTGTTCGTGGTGTGAA, TTCAACACCACGAACACGCC,  
TTAGTATTACGGATAGCGGT, ACCGCTATCCGTAACTAAA,  
GTCGTTGTTCGTGGTGTGAA, TTCAACACCACGAACACGAC,

ATATGTAAAT, ATTTACATAT, ATTTGTATAT, ATATACAAAT,  
TTATGTAAAT, ATTTACATAA, ATTTGTATAA, TTATACAAAT,

GAATATTTA, TAAATATTC, TGAATATTT, AAATATTCA,  
GAATATGTA, TACATATTC, TGTATATTT, AAATATACA,

ATAAT, ATTAT, ATTAT, ATAAT,  
GTAAT, ATTAC, ATTAT, ATAAT,

AATGTAAAT, ATTTACATT, ATTTGTATT, AATACAAAT,

ATTTGTATATT, AATATACAAAT, GGTATGTAAAT, ATTTACATACC,  
ATTTGTATATT, AATATACAAAT, AATATGTAAAT, ATTTACATATT,  
ATTTGTATATT, AATATACAAAT, AGTATGTAAAT, ATTTACATACT,  
ATTTGTATATT, AATATACAAAT, GATATGTAAAT, ATTTACATATC,

AGGAGT, ACTCCT, ATTTTT, AAAAAT,  
GGGAGT, ACTCCC, ATTTTT, AAAAAT,  
GGATATGTTGGGTATGTT, AAACATACCGAACATATCC,  
GGATATGTTGGGTATGTT, AAACATACCGAACATATCC,  
GGATATGTTGGGTATGTT, AAACATACCGAACATATCC,  
AGATATGTTGGGTATGTT, AAACATACCGAACATATCT,  
TCGTTCGTTTAGATAT, ATATCTAAACGAAACGA,  
ATATTAGAGCGGAACGG, CCGTTCCGCTCTAAATAT,

CGTTACGGTT, AACCGTAACG, AATCGTGACG, CGTCACGATT,  
CGTTACGGTT, AACCGTAACG, GATCGTGACG, CGTCACGATC,  
CGTTACGTT, AAACGTAACG, AAGCGTGACG, CGTCACGCTT,  
CGTTACGTT, AAACGTAACG, GAGCGTGACG, CGTCACGCTC,

TTTACGTATGA, TCATACGTAAA, TTATGCGTGAA, TTCACGCATAA,  
TTTACGTTTGA, TCAAACGTAAA, TTAAGCGTGAA, TTCACGCTTAA,  
TTTACGTTTA, TAAAACGTAAA, TGAAGCGTGAA, TTCACGCTTC,  
TTTACGTATTA, TAATACGTAAA, TGATGCGTGAA, TTCACGCATCA,  
  
AATTAAATTAA, TTAATTAAATT, TTGATTGATT, AATCAATCAA,  
TATTAATTAA, TTAATTAAATA, TTGATTGATG, CATCAATCAA,  
  
TAATTAT, ATAATTA, ATGATTG, CAATCAT,  
  
TAGGTAA, TAACCTA, TGATTAA, TAAATCA,  
  
TTTTAAATATTTT, AAAAATATTTAAA, GGGGGTGTGTTGGGG,  
CCCCAAACACCCCC,  
TTTTAAATTATTTT, AAAATAATTTAAA, GGGGTGGTTGGGG,  
CCCCAAACCCACCCC,  
TTTTAAATTTTTT, AAAAAAATTTAAA, GGGGGGGTTGGGG,  
CCCCAAACCCCCCCC,  
TTTTAAATAATTTT, AAAATTATTTAAA, GGGGTTGTGTTGGGG,  
CCCCAAACAAACCC,  
  
GAGGCGGGGG, CCCCACCTC, TTTCGTTT, AAAACGAAA,  
GAGGTAGGG, CCCTACCTC, TTTTGTGTTT, AAAACAAAA,  
AAGGCGGGGG, CCCCACCTT, TTTCGTTT, AAAACGAAA,  
AAGGTAGGG, CCCTACCTT, TTTTGTGTTT, AAAACAAAA,  
  
GGGGGCGGGGT, ACCCCGCCCCC, ATTTCGTTTT, AAAAACGAAAT,  
GGGGGCGGGGT, ACCCCGCCCCC, GTTTCGTTTT, AAAAACGAAAC,  
TATTATTTAT, ATAAAATAATA, GTGGGGTGATA, TATCACCCCCAC,  
GATTATTTAT, ATAAAATAATC, GTGGGGTGATT, AATCACCCCCAC,  
  
ATTACGTGAT, ATCACGTAAT, ATTACGTGAT, ATCACGTAAT,  
ATTACGTGAT, ATCACGTAAT, GTTACGTGAT, ATCACGTAAC,  
  
TTTTATATGG, CCATATAAAA, TTATATAAGG, CCTTATATAA,  
TTATATATGG, CCATATATAA, TTATATATGG, CCATATATAA,  
AAATAAT, ATTATTT, GTTGTGTT, AAACAAC,  
AAATTAA, TTAATT, TTAGTTT, AACTAA,  
AAATTAT, ATAATT, GTAGTTT, AACTAC,  
AAATAAA, TTTATTT, TTTGTGTT, AACAAAA,  
  
ATTTTCGGAAATG, CATTCCGAAAAAT, TATTTTCGGGAAAT,  
ATTTCCCGAAAATA,  
ATTTTCGGAAATG, CATTCCGAAAAAT, TATTTTCGGGAAAT,  
ATTTCCCGAAAATA,  
ATTTTCGGGAAATG, CATTCCCGAAAAT, TATTTTCGGGAAAT,  
ATTTCCGAAAATA,  
ATTTTCGGGAAGTG, CACTTCCGAAAAT, TATTTTCGGGAAAT,

ATTTCCGAAAAATA,

AATAGATGTT, AACATCTATT, AATATTGTT, AACAAATATT,  
AATAGATGGT, ACCATCTATT, ATTATTGTT, AACAAATAAT,

GTATAAATA, TATTTATAC, TATTTATAT, ATATAAATA,  
GTATAAATG, CATTTATAC, TATTTATAT, ATATAAATA,  
GTATAAAA, TTTTTATAC, TTTTTATAT, ATATAAAA,  
GTATAAAAG, CTTTTATAC, TTTTTATAT, ATATAAAA,  
TTATAAATA, TATTTATAA, TATTTATAG, CTATAAATA,  
TTATAAATG, CATTTATAA, TATTTATAG, CTATAAATA,  
TTATAAAA, TTTTTATAA, TTTTTATAG, CTATAAAA,  
TTATAAAAG, CTTTTATAA, TTTTTATAG, CTATAAAA,  
GGGGGTTGACGTA, TACGTCAACCCCC, TGCGTTAACCTT,  
AAAAATTAACGCA,  
GGGGGTTGACGTA, TACGTCAACCCCC, TACGTAAATTTC,  
AAAAATTAACGTA,

TGACGTATATTTT, AAAAATATACGTCA, GGGGATATGCGTTA,  
TAACGCATATCCCC,  
TGACGTATATTTT, AAAAATATACGTCA, GGGGATATGCGTTA,  
TAACGCATACCCCC,  
ATGATTTAGTA, TACTAAATCAT, TGTTGAGTTAT, ATAACCTAACCA,  
GTTAT, ATAAC, ATGAT, ATCAT,

TTACGTGA, TCACGTAA, TTACGTGG, CCACGTAA,  
TTACGTGG, CCACGTAA, TTACGTGG, CCACGTAA,  
TTACGTGG, CCACGTAA, TTACGTGA, TCACGTAA,  
TTACGTGA, TCACGTAA, TTACGTGA, TCACGTAA,  
GACGTT, AACGTC, AGCGTT, AACGCT,

TGACGTGT, ACACGTCA, ATACGTAA, TAACGTAT,  
TGACGTGG, CCACGTCA, TTACGTAA, TAACGTAA,  
CGGTTATTTG, CAAAATAACCG, TAAGATGGTCG oder CGACCATCTTA

enthält, welche zu einer solchen DNA komplementär ist oder dieser entspricht, wie sie entstehen würde, wenn ein ebenso langes DNA Fragment, welches über seine Sequenz oder Sekundärstruktur die spezifische Lokalisierung von Genom/Chromatinabschnitten innerhalb des Zellkerns herbeiführen kann, einer chemischen Behandlung gemäß Anspruch 1 unterzogen würde.

19. Verfahren nach einem der Ansprüche 16 bis 18, dadurch gekennzeichnet, daß die zur Amplifikation verwendeten Oligonukleotide außer den in den Ansprüchen 16 bis 18 definierten Konsensussequenzen mehrere Positionen enthalten, an denen

entweder irgendeine der drei Basen G, A und T oder irgendeine der Basen C, A und T vorhanden sein kann.

- 20.Verfahren nach Anspruch 19, dadurch gekennzeichnet, daß die zur Amplifikation verwendeten Oligonukleotide außer einer der in den Anspruch 18 beschriebenen Konsensussequenzen nur maximal zusätzlich so viele weitere Basen enthalten, wie es zur gleichzeitigen Amplifikation von mehr als einhundert verschiedenen Fragmenten pro Reaktion aus der chemisch behandelten DNA, berechnet entsprechend Anspruch 8, erforderlich ist.
- 21.Verfahren nach einem der vorangehenden Ansprüche dadurch gekennzeichnet, daß die Untersuchung des Sequenzkontextes aller oder eines Teils der in den amplifizierten Fragmenten enthaltenen CpG Dinukleotide oder CpNpG Trinukleotide gemäß Anspruch 1 c) durch Hybridisierung der bereits in der Amplifikation mit einem Fluoreszenzmarker versehenen Fragmente an einen Oligonukleotid- Array (DNA Chip) erfolgt.
- 22.Verfahren nach einem der Ansprüche 1 bis 20 dadurch gekennzeichnet, daß die amplifizierten Fragmente auf einer Oberfläche immobilisiert und anschließend eine Hybridisierung mit einer kombinatorischen Bibliothek von unterscheidbaren Oligonukleotid- oder PNA-Oligomer-Sonden durchgeführt wird.
- 23.Verfahren nach Anspruch 22 dadurch gekennzeichnet, daß die Sonden mittels Matrix-assistierter Laser-Desorptions/Ionisations Massenspektrometrie (MALDI-MS) anhand ihrer eindeutigen Masse nachgewiesen werden und damit der Sequenzkontext aller oder eines Teils der in den amplifizierten Fragmenten enthaltenen CpG Dinukleotide oder CpNpG Trinukleotide entschlüsselt wird.
- 24.Verfahren nach einem der vorangehenden Ansprüche, dadurch gekennzeichnet, daß die Amplifikation wie beschrieben in Schritt b) des Anspruchs 1 durch eine Polymerase Kettenreaktion ausgeführt wird, in der die Größe der amplifizierten Fragmente mittels auf weniger als 30 s verkürzter Kettenverlängerungsschritte

begrenzt wird.

25.Verfahren nach einem der vorangehenden Ansprüche, dadurch gekennzeichnet, daß nach der Amplifikation gemäß Schritt b) des Anspruchs 1 die Produkte durch Gelelektrophorese aufgetrennt werden und die Fragmente, die kleiner als 2000 Basenpaare oder kleiner als ein beliebiger Grenzwert unterhalb von 2000 Basenpaaren sind, durch Ausschneiden von den anderen Produkten der Amplifikation vor der Auswertung gemäß Schritt c) des Anspruchs 1 abgetrennt werden.

26.Verfahren nach Anspruch 25 dadurch gekennzeichnet, daß nach der Abtrennung von Amplifikaten bestimmter Größe diese vor der Durchführung des Schrittes c) des Anspruchs 1 nochmals amplifiziert werden.

27.Kit, enthaltend mindestens zwei Primerpaare, Reagenzien und Hilfsstoffe für die Amplifikation und/oder Reagenzien und Hilfsmittel für die chemische Behandlung gemäß Anspruch 1 a) und/oder eine kombinatorische Sondenbibliothek und/oder einen Oligonukleotid-Array (DNA-Chip), soweit sie für die Durchführung des erfindungsgemäßen Verfahrens erforderlich oder dienlich sind.

## SEQUENZPROTOKOLL

## ALLGEMEINE ANGABEN:

## ANMELDER:

NAME: Epigenomics AG  
STRASSE: Kastanienallee 24  
LAND: Berlin  
POSTLEITZAHL: 10435  
TELEFON: 030-243450  
TELEFAX: 030-24345555

BEZEICHNUNG DER ERFINDUNG: Verfahren zur parallelen Detektion  
des Methylierungszustandes von  
genomischer DNA

ANZAHL DER SEQUENZEN: 31

## COMPUTERLESBARE FASSUNG:

DATENTRÄGER: Diskette  
COMPUTER: IBM PC-kompatibel  
BETRIEBSSYSTEM: PC-DOS/MS-DOS

## DATEN DER JETZIGEN ANMELDUNG:

ANMELDENUMMER: nicht bekannt  
ANMELDETAG: 6.12.2000

## ANGABEN ZU SEQ ID-NO:1:

## SEQUENZCHARAKTERISTIKA:

LÄNGE: 12 Basen  
ART: Nukleinsäure  
STRANGFORM: Einzelstrang  
TOPOLOGIE: linear

ART DES MOLEKÜLS: chemisch vorbehandelte Genom-DNA

SEQUENZBESCHREIBUNG: SEQ ID-NO:1:

AGTAGTAGTA GT

12

## ANGABEN ZU SEQ ID-NO:2:

## SEQUENZCHARAKTERISTIKA:

LÄNGE: 12 Basen  
ART: Nukleinsäure  
STRANGFORM: Einzelstrang  
TOPOLOGIE: linear

ART DES MOLEKÜLS: chemisch vorbehandelte Genom-DNA

SEQUENZBESCHREIBUNG: SEQ ID-NO:2:

AAAACAAAAAA CC

12

## ANGABEN ZU SEQ ID-NO:3:

## SEQUENZCHARAKTERISTIKA:

LÄNGE: 973 Basen  
ART: Nukleinsäure  
STRANGFORM: Einzelstrang  
TOPOLOGIE: linear

ART DES MOLEKÜLS: chemisch vorbehandelte Genom-DNA

SEQUENZBESCHREIBUNG: SEQ ID-NO:3:

AGTAGTAGTA	GTAGCGTTT	TGAAGTTTT	TGTGGAAGGT	GAGAAATTAA	TCGATAAGTT	60
TTTTAGTAGG	AGTGTTTT'G	GGGAGGGAGT	GAGTGGGAGA	TTAGAAATGG	GGTCGGTGGA	120
ATATTTTTGT	AAAATTTAG	GAATTATTAG	TATTTTATT	TTTTTTATAA	AGATTGTTT	180
TATGTTTGAG	TTGTTTTATA	TACTGTAGAA	ATTGAGAATG	AAACAGTGT	AACGATCGGT	240
GTGATTATT	GAAAGGCAT	GCGCGTTT	TTTTGTTAGG	TAGAAATTGTT	TTAGGAAGTA	300
TGAAAGTGA	CGTAGTCGG	TGTTAGGTGT	TAGCGATTG	AGTTCGTTT	CGGGTTTTAT	360
TTATTTTTAG	TTCGGTTTT	AGATTTTT	CGAGGC GTTT	TTTTTTGTC	GTTCGATT	420
TGAGCGGAGC	GTTCGGGGT	GTGAGGAGAA	TCGGTAAATT	TTCGCGGGCG	TTGGGCGTCG	480
TAGAGTCGTC	GCGCGTTAG	TTTCGTTACG	TTGGTTGTCG	AGCGTATTTC	GGTTGCGTTC	540
GGCGGGGAGG	CGTCGAGGGT	AGTTAAGGGG	AGTAGGTTAC	GTGAGGAGGAG	GAGTTGATT	600
TATTTTTAG	GCGGTAGGCG	TATGCGTATT	TTTATTGTC	GTTCGGTTC	GGAGGTTTAC	660
GTGGAATCGT	ACGTGTTGG	TTTGTTAGTT	AGGGGTTTT	GGTCGGGGC	GCGTAGGGGC	720
GGGTTCTAG	TGGGATTTCG	GGAGAGGGGC	CGGGCGGGC	GGAGCGTTG	GAGATTAGT	780
TTGCGGGTTT	CGTAATTATT	ATTCTGTATA	ATTATAGTTT	TTGGAGAGTT	GTTCAGGTTT	840
TTTGCGGGGT	TTTTTACGAA	TTTATTATAT	TTAATATTG	TGAATTAGA	AATTAAGATA	900
AAGATTGTAT	TTTGTTTTG	AAATTATTA	TTTCGTAAT	TTAGAATTAA	TTTGTAAT	960
GGGTTTTGTT	TTT					973

ANGABEN ZU SEQ ID-NO:4:

SEQUENZCHARAKTERISTIKA:

LÄNGE: 1890 Basen

ART: Nukleinsäure

STRANGFORM: Einzelstrang

TOPOLOGIE: linear

ART DES MOLEKÜLS: chemisch vorbehandelte Genom-DNA

SEQUENZBESCHREIBUNG: SEQ ID-NO:4:

AGTAGTAGTA	GTAAAGTTTA	AGAGTGTAT	TTTTTAGAT	ATAGTATAAT	ATTGTTTTG	60
TAAAAGGTTA	TTTATATGAA	TATAATGTT	TTTGGGTTAA	TTAACCGTT	TTGAGAAATAG	120
TTAGGTTAA	ATTTTAGGT	TTTTTATT	GGGTTATTAA	AGTTAGGATC	GCGACGTGAG	180
TTCGGGGTGA	GTCGAGGGGT	ATTCGGGTC	GAGGTTATT	GAATCGAAGT	TATGAGGC	240
GGGTCGGTAA	TCGGAAGTGG	TTAGGGAGA	GTTGTACGAG	ATTGGGGGT	TGTGATTG	300
AAATAAAATA	AAATAAAATA	AAATTAAAT	TGTTTGGGA	TTGATTTTA	AAAGAGCGTT	360
TTTTGGTTT	AAGAGGTCG	CGTCGCGGA	GATGCGTT	TAAAGGGTTG	TTTTTAAGC	420
GTGTAGGTCG	CGTACGGGTT	TTTTAGCGG	CGGGTAAAAA	TGGGCGTCG	TATTGGGAG	480
GCGTTTGTTT	AGCGTTCG	GCGTCGTTA	TAGAGTACGT	TCGTTTGCGG	TTTAGAGCG	540
TCGTTTTTT	GTCGTTTCG	TCGTCGTT	TGACGTTGCG	ATCGCGTCG	GTTATCGTT	600
TCGTTTCA	CGGTTACGTT	TGTTTAAAT	CGCCGCGGCG	TTTTTTAGGT	GTTGTTGGC	660
GGGTTTCGTT	TTCGTGTTT	AAGGTTCGT	TTCGCGCGT	AGTCGCGC	TCGTTGTTT	720
TTTTTCGTT	TTATAGTTTC	GTTTTATAG	TTTCGTTCGT	TTTTTAAGCT	TCGTTTTTA	780
GGAATTTCGCG	CGTCGAAGGT	TAGGTTGGG	CGGAGCGTAT	AGCGTTGGC	GTTGGGGAGG	840
TTGCGTCGTA	GTATTCGTT	GGTTAGGATT	AAGTGGGTC	GAGGCGGACG	TGAGAAGGGT	900
CGGGTTAAAGA	TGGCGGTGTA	GGTGGTGTAG	GCGGTGTAGG	CGGTTTATT	CGAGTTTGAC	960
GTTTTTTCG	TTTGTTTAA	TTACGTTTG	AGTATAGAGA	AGGAGGAAGT	AATGGGGTTG	1020
TGTATAGGGG	AGGTGAGTAG	GTTTGTAGT	TTGGATGGAA	TTTTGTTGAG	TAGTTTGT	1080
GTGTTTTCGG	GGTGGGTGTC	GGTAGTTTT	AGGGTTGCGG	AGGTTATAGG	TATTTTCGAT	1140
TTAGGTTTT	GGATATT	TATATAGTT	TGTCGCGGG	AGTTTCGTT	TTTTGGTCG	1200
TTTGATATT	GTTAGTTT	TTGCGAGTT	TCGGCGTTT	GTATAGGTT	CGGGAGTTT	1260
GTTTTTTTT	TTCTAGTTT	GGGTTTTT	TTAGATTG	GTTCGTT	CGGTTGATTA	1320
AAATTGTAAG	GTAGGTTAGA	AAGATATTG	AGTATAATGA	GGATGTTCGG	GTATTACGAC	1380
GTAGTAGTTG	ATTATAGTTA	GAGTTTTT	TTTCGTTTCG	AGTTTTTGT	TTAGGGAGTA	1440
GAGATATTAA	TTAAAGTATT	TGAAGGGTAT	CGAAGAGTT	AATAGAAGGT	GCGGGGTTT	1500
AAGGAAGTAA	AAAGTTTCGT	TGTATTG	TGAGGAGGG	TGAAAGAGGA	TGAGGAAATA	1560
TAGTTAGTT	GTTTATAGTT	AAAGTAATT	TTTAGTTT	TTATATTATG	TGCGTGAATA	1620
TATGATTAA	TTGTTATATA	ATTTGTATT	ATATATGTTA	AATAAACGTA	ATGTGATTAA	1680

TAATGTTTT TGTCCCCCCC TATTTAGTT GAACGATGAT ATAAGGTAAG ATTGTATTTG	1740
TTTATTATA TTTTATAATT TTTAAGTAT TGTAAATTAA AAATTATGGC GATTTGGTA	1800
GTATTGTGTT CGGATCGTGT TTTAAATAA TTATTATTAT TATTCGGGG TTTTTTAAT	1860
GATATAATT TGAAATTGTT TTTTGTTTT	1890

ANGABEN ZU SEQ ID-NO:5:

SEQUENZCHARAKTERISTIKA:

LÄNGE: 2222 Basen

ART: Nukleinsäure

STRANGFORM: Einzelstrang

TOPOLOGIE: linear

ART DES MOLEKÜLS: chemisch vorbehandelte Genom-DNA

SEQUENZBESCHREIBUNG: SEQ ID-NO:5:

AGTAGTAGTA GTTTGGAGA GTAAGTGGGT TTTATGTAAA GTGTTATTTG TTATATATTG	60
TTATTTTAN AGGATTTAG AATATTTCG AATTGAAAAA TAGAGTGAGT GTGGAGGGGA	120
GGGGGGTGT TACGTGGAGG AGGGTCGCGA GAAGGAGAGT GCGTCGGGTT GATGGTTAGC	180
GGTGTTCGG GGGTCGATGT CGGCGGGAT CGGCCGGCGC GGGGTGGGGC GACGGCGGGG	240
CGCGCGTCGA GTAGTAGAGG GGCGTTTCGT AGAGGTCGGG GGGGGCGCGG GTGTTCGCGT	300
CGTTTTGTTA GTAGTCGTT CGTCGTTTT TTCGTTTCGT TCGTAATCGT CGCGGGATT	360
GTCGGTCGG GTTTTGGAA TTTTTACGG GGGATAGTCG ATCGGGGGCG TCAGGGGTTCG	420
TTTCGTATT CGATTCGGAC GGCAGTAGTA GGGGGAGGGG TGGTGTATCG AGGAAGTTAA	480
GGTTTATTAA TTTGTTCCG CGGGTCGGCG GTTCGTTGGT TCGTTGGTC GGAGAAGTGT	540
TGCGTTAGG TTGGTTCGTA GGAAACGGCG GCGGCGGTTT ANTTNTANTT TTNNNNNTNT	600
TNNTTGNAN TNNTTTTTA TAGGGGAGGG GTAGGCGGTT CGCGGGTTTC GCGTCGGTCG	660
GGCATTGGG TACCGCAGGG AGCGGGTAGG GTAGGGGAA ATAAATTAAG GTCGAGATT	720
AGAGTCGGAC AGCCGGGGAG GAGTAGCGGC GAGAGGTAGG AGAGGTAGAG AAGAAGAAAG	780
GAGGGAGAGA GGGGCGAAG ACGGTCGGGA AATTCCGGGT TGTAGTTTCG CGTCGCGTCG	840
GAGTCGTGAC GGATTATTG TAGCGGTGCG GTTGATTTCGT AATTAAAGG TACGAAAAAA	900
GGGGAGGGTA GAGAGGGAGG GGAAAGCGGA GTGTTGTAGC GTCGGGGCAG GGGCGGGTTT	960
TCGCGACCGT CGTATATTGCG GGAATTGTN GTTTCGTTGC GGGCGAGTGT CGCGTGT	1020
GGCGAGTTT GTGTTGGGAA NTTTGAGCGC CGGGATTACG TTCGTTTTT AGTCGGTCGT	1080
TTTTGCGTG TAGGGTGGGG GGAGGTTAGG GGCGGGGGCG CGGACGTCGG TGACGTCGCG	1140
TGGCGGAGTT TTTCGGTTA TGTGGTTGGA GGCAGGGCGGG GAAGAAGGTA AGGTTGGAG	1200
GGGAAGGAGG AAATGCGAGG GTTTTCGGC GGGAGGAAGC GCGAGGGGTG CGCGAGTGT	1260
AGCGAAAGGG GGATAGGGGT TTTGGAGTGG AAGTTTGCG GGAGAGGAGG AGTAGAGGAG	1320
TAGTTTTAGG GTGTTAGAAT TATTGCGATG TCGTATTAAA AAATAGGAA AATTAAAGTT	1380
TTGCGAGAGG TAGTTAAAGG TATTGTTGTA TTCGTTGATC GATTGGATT TATTATAGAA	1440
TTGTACGGTT TATTAGGATT GTTGTATT TCGGTAGTA TTCGTTAGTT TATTGTAAT	1500
ATGAATTGGT AGAGTTAGGA ATTAGGTTGT AAAGATAGAG AAGCGAGTTA AAGTTGGTCG	1560
TAGTCGAGG CGTGGGGAGA ATTGGGAAAC CGGAGGAGAA CGGATATT TTTATTGTTAG	1620
AAAGATTTT TATTAGTT TGTATTAA TAAATCGTT AGATTTTTT TTGGCGGAGT	1680
TTATTAGTT TTGTTAAAAA AAAGAAAAAA ATTCAATGT AGATTGTTCG TTTATTTTT	1740
TAGTAGAGTT ATATTATT TGTGGTAAT TATTGTTA GAAAATTAG ATTATAATAG	1800
GAAATTATAT TTAGAAAAAGT ATAAGGAGGA AATTATGTT CGAGAGAAGA AATAAAGTTG	1860
TTAGGAGAGG TGTGATGAGG ATAACGAAGA AAATATTG TAATTATT TATTAGTTAT	1920
TTTTTATAA GTTTGATAA TCGTCGTAT TCGTGTGTGTT GAGAGAGTGT GTGTGTGT	1980
GTGTGTGTGT GTGAGAGAGA GAGAGAGGAT GTTTTGAG TGATCGGAAA TTTTTATTG	2040
ATAAAAGTTT TGAGTTATT TATAAAAAGT TATTGTTG TTTTTTTA AAAAAGTTT	2100
AATGATTTT ATATGAGAAT TTTATTTTT TTATGTTA ATTATTTTT TATATGATGA	2160
GTATGTTATT TTTTTAATA AACGTTAGC GATTGTTTT AGGATTAAT GGTTTTGTT	2220
TT	2222

ANGABEN ZU SEQ ID-NO:6:

SEQUENZCHARAKTERISTIKA:

LÄNGE: 307 Basen

ART: Nukleinsäure

STRANGFORM: Einzelstrang

TOPOLOGIE: linear

ART DES MOLEKÜLS: chemisch vorbehandelte Genom-DNA

SEQUENZBESCHREIBUNG: SEQ ID-NO:6:

AGTAGTAGTA GTTTCGTTAC	GAATACGATT AGTTTATATT	GGTTGTTAG TATTCGGTT	60
AGTTTTGTG GGACGTTGGC	GCGTATATCG TTTCGTCGGG	GGGTCGGGA TTTCGGGGAT	120
TTCCGGGTCG AGGACGAGGG	AGGCAGTAG TCAGCTGGTT	TCGACGATT TGCGATCGAG	180
GCGTTTGTAG TTGTTCGGGG	CGAGGCGGGC GATTTAGGT	TTTCGAGTAT TTTCGAGATA	240
TTCGCGTAA GTTTGGTTTC	GGGGTGGTT GTCGTCGGT	TTCGTTGCG TTTTGGTT	300
TTGTTTT			307

ANGABEN ZU SEQ ID-NO:7:

SEQUENZCHARAKTERISTIKA:

LÄNGE: 523 Basen

ART: Nukleinsäure

STRANGFORM: Einzelstrang

TOPOLOGIE: linear

ART DES MOLEKÜLS: chemisch vorbehandelte Genom-DNA

SEQUENZBESCHREIBUNG: SEQ ID-NO:7:

AGTAGTAGTA GTTGTGTTG TTCAAGGTTT CGGTGTTAGTT	GGAAAGTTTG GAGGTGGGGA	60
GAGGGATGTT AGGTAAGTGG TACGGCGAGC	GTAAGGGAAG GGGTTAGTTA TTGATTAGCG	120
GTAGTAATTG TAGGAATCGT CGTCGTTAGTT	GTAGTCGTTT TTTCGTTCTG TTTTCGGGT	180
TTTCGGAAA ATGGTTGTGG GGTTGGTCGC	GTCGTTAAGT TTGTTTTCGC GCGGTGAAGA	240
GCGGTTGTT TGCGGGAGTC	GTTCGTTAAAT TTCCGCGGC	300
TGGTAGGATT CGAGGGGTTA CGAGTTGATA	TTTTTTGGG TTGTATAAAA AGTTGAGGCG	360
GGCGTTGGGA GGAGGTAGCG	GTGTTGCGG TGCGGTTTT	420
CGGGTTATTT TTTCGTTCTG	TTTCGTTTT TTTCGTTTACG CGGGTTTTTC	480
GGGGTTTGG CGGTTTCGG TTGAAGGTCG	CGGTTTTGT TTT	523

ANGABEN ZU SEQ ID-NO:8:

SEQUENZCHARAKTERISTIKA:

LÄNGE: 653 Basen

ART: Nukleinsäure

STRANGFORM: Einzelstrang

TOPOLOGIE: linear

ART DES MOLEKÜLS: chemisch vorbehandelte Genom-DNA

SEQUENZBESCHREIBUNG: SEQ ID-NO:8:

AGTAGTAGTA GTAGCGTCGG GAGTCGCGT	AGAAACGATT TGATTTTCG CGGCCATT	60
TTTTTTGCG ATTGGCGTT CGCGGAGAAA	ATTAGTTG TCAGATTTTC GTTTATTATT	120
TCGTTTTAT TTATCGCGC	GTTCGTTGGG AATTGTTAGTT TTTCGTCAG CGGTGGTAG	180
CGGCAGGTC	GTCGGCGGAT TTAGCGTTT AGGAAGTTT GCGTAGGCGT AAGCGTTTT	240
TAGGTAATTA CGTCGCGGTT	CGAGGTTGTA GATTTGGTA GGGGATATAA TTGGAGAATG	300
AGGGTGGTCG	GGTGGGAGTA GGGATTAGGT TTATTTAGA GTTCGGAGTT TGGAGTTCGT	360
ACGATTGAG	GTTCGTTTT TTTTTGTTT TATAGAAAGA GCGGAAAGTT TAAGAATCGG	420
GTCGCGTTTG	GTGAGTTG ATAATGTTT GGTTTACGC GTATTGACGT CGATTTGT	480
ATGTTAGTG	TCGTTGCGGG GTTGGTATTG CGGTTGGGT TTGGCGTGA GGAGTTGGT	540
TTAGTCGTT	TTGTTTTTC GCGGAGCGTT TGGTCGTTGG TAAGTTAGG TTAGGTTGTT	600
TCGGGGACGT	AGGGTCGCGT AGACGTTTT TTACGTTCG GGGTTTTGT TTT	653

ANGABEN ZU SEQ ID-NO:9:

SEQUENZCHARAKTERISTIKA:

LÄNGE: 1461 Basen  
 ART: Nukleinsäure  
 STRANGFORM: Einzelstrang  
 TOPOLOGIE: linear

ART DES MOLEKÜLS: chemisch vorbehandelte Genom-DNA

SEQUENZBESCHREIBUNG: SEQ ID-NO:9:

AGTAGTAGTA	GTAGGGCGGTG	GTAGGTCGG	GTCGGGGCGGT	TATTAGCGGT	TTTCGGAGGT	60
ATTTAGGTTT	TACGGTCGAG	TTTTGGTCG	GAGATCGGTC	GTTCACGGGG	GGTTATAGTA	120
CGCGTCGTT	GTTAGGATG	GAGAGGCCGG	TTTAGGTTTC	GGTCGGAGC	GAGTTTTTA	180
GGGTTTGTCG	ATACGGCGGG	GTTCGGTGGT	CGGTATTG	TTCGTACGTG	TCGAGGGGT	240
TTTCGGGTTT	TCGGTATTAT	GGTTATTATC	GGGGTTTCGA	TTACGACGAG	GTCGATGGTT	300
CGGGTAGCGG	GGGCGGCGAG	GAGGTTATGG	TCGGGGTTTA	CGACGCGTTA	TTTTTCGTAC	360
GATACCGGTT	TCGGGGCGTT	ATCGGGCGTT	CGTTAGGAT	TTTCGGGTT	TCGGGGTCGG	420
TTTGCGTTTC	GTTTTTCGG	TACGGTCGGC	GATTTTTAA	CGGTTATTAT	TCGGCGTACG	480
GATTGGTTAG	GTTCGCGGG	TCGGGTTTA	GGAGGGTTT	GTACGAATT	TATAGCGAGA	540
GTGACGATGA	TTGGTGTAA	GTTCGGGCGA	GGTGGCGTTC	GTTCGGTTT	TTACGTATTT	600
TACGTATATA	TTTATTTCGA	GGAGTCGCGT	AGAGGTCGCG	GGGGTTTAGT	ATAGAGGGTT	660
CGGGAGAGGG	TTAGTCGGGA	GATTTAGAT	TTTGAGAGG	TTAGGGTTGG	GTTATAAGGG	720
TGTTTCGTAG	AGATTTTCGG	TTAAAAGAGA	TTTTTTGGG	TAGTTACGGC	GTTCGGTTAAT	780
TAGTTTCGAT	TTTTTATTT	ACGATAGGGG	TTTCGGGTG	GGAGGTAGGG	AGTAGATAAA	840
TTATATAGTT	AAGGGATTTG	AATTAATT	GTTATTTTG	GAGAATT	GGGAATATGA	900
AAAAAAA	AAAAAAA	AAAAAAATA	TTTTAAAAG	AAAAAACGGG	GAGAAAAAAA	960
TAGTTTTAT	TGATGAGTT	TATTATTTA	ATTGAATT	TTTTTTTTT	GATGAAGATA	1020
GTGCGGGTC	GAGTGCAGTA	AAGAAGTTAG	AAGGAATTAG	AATTTAGT	TTTTATATTT	1080
ATTATTAGAT	ATATTATAT	TTATATACGT	TTTAGATAT	ATATAAGAGT	GTTCGCGGT	1140
TATATTAAAT	TTTATTATTA	TTGTTGTAG	AAATTAATT	AAAAAAATAA	TAATAATAAT	1200
AAATAATT	AAAAGGATA	AAAAAATTAA	TGATTGAGAA	AAGAGGTATT	TTTTTTGAT	1260
ATTTGGTTT	GTTGAAATA	ATAAAAGAAG	AAGAAAATT	TATTATTATT	ATCGATTTT	1320
TTGTTTTT	TTTTTTTTT	TTTATTTTG	TTTGAACATC	GTGGGTTGG	GATTGTGAAT	1380
TATTGTATGA	TATTA	AAAAA	ATAAAAAAA	GTTGAATTAA	AGGGTTTTG	1440
GATAGGAGTG	GT	TTTGT	T			1461

ANGABEN ZU SEQ ID-NO:10:

SEQUENZCHARAKTERISTIKA:

LÄNGE: 2536 Basen  
 ART: Nukleinsäure  
 STRANGFORM: Einzelstrang  
 TOPOLOGIE: linear

ART DES MOLEKÜLS: chemisch vorbehandelte Genom-DNA

SEQUENZBESCHREIBUNG: SEQ ID-NO:10:

AGTAGTAGTA	GTAATCGGGT	CGTCGTCGTC	GCGGGGTTAA	GAGATTACGT	TTAGTAAAAA	60
TTTTTTAGG	AAAGAGGTTC	GAGAGTTTT	TTTTTATTT	TTAAGGGTG	TTGATTGTTG	120
CGCGTTACGT	TACGTATTAA	AAGGTTCGCG	AAGAAAAACG	CGGGCGAAGT	TTAGTATT	180
CGAATAGAAA	GGGAATT	TTAGAAGGG	GGGTTACGG	TTTGAGGGT	TTTGGGTGG	240
GGAAAGTAGT	ATTGGAGGG	GAAGATAGAA	GGGATGGTAA	GATAGAGGAA	ATTACGGGG	300
TAAAAGGAA	GGGATGAGGA	GT	TATAAAAGGG	ATTGACGGAC	GGGATGTAGA	360
GTAATGGTT	TTTTCGAA	AATGGATTG	TAGGTTTTT	CGTATTTTA	AATAATT	420
TATTATTGA	AAGCGGGGAG	TTTGAGTTG	GGGTTGCGTA	GATTAAAGTA	GGGTAAGACG	480
AAGGAAAGA	GGGAGGGGAG	TTTGAAGTA	AAGTGAATAA	AGAGGTAAA	GAGGTAAAGT	540
TGGAAATGGA	AACGAGAAAG	GGAGTTGGA	AGGGATCGTT	TTGGAAGGGT	TTTGAAAAG	600
GTTTAGAAT	GGGATGTTGG	GGAGGTAAAG	GGGGATAGTT	TCGAGGAAGG	GTATT	660
AATGTAAAGG	AGCGTTTAG	TTTTGGGAA	GAGGAAGGG	TGGGAGGGCG	GGGAGGTAGT	720
CGAGGTTCGG	GGAGTGGAA	AGAGGGGTT	TTGCGCGGGG	TTAGGAGGG	GAGGTAGG	780
TCGGAGGGAG	GCGATTATCG	CGGAAAGAGT	TAGAGGATGG	GAGGGAGGG	TGTTGCGGTT	840
GAAGAGGGGA	TGTGGATT	TTAATGT	GGGAGAAGAG	GGAGGTGAAG	CGAGATAAAA	900

GGAAAGGGGC	GAATAGGGAG	AAGAGGAAGG	AGTTTGCAC	GCGATGGGG	AGGGGAGGG	960
GGTCGTGGG	AGCGGTCGGG	GGCGCGGGG	ATGGGAAGGG	GTCGGGCGG	C GGCGTGGT	1020
ATTTTAGGTT	GCGGATTTT	TTTTGGGG	GCGGCCTGT	CGGC GGCGG	GTT CGTAAT	1080
TTTCGGTT	TTTCGTTCGT	TTTTCGTT	TTTCGGCG	CGGGCGGGG	T CGGTACGGT	1140
ACGGTTATA	GCGGCGGTT	TTTGCGTCG	AGTTTGGAT	GTTGTTTGT	GGGAGGCGG	1200
GGTAGCGGT	GCGGTAGCGG	TTCGGTTCGT	ACGGTTATT	TCGTTCGCG	TAGTAGGC	1260
TCGCGGGTCG	AGTTTTTAG	TAGTCGGCGT	CGGC GTCGC	GGCGGCGGG	GCGGCGGTT	1320
TGGTTTTTG	CGTTTGT	TTTCGTTTC	GGCGTTTG	GGCGGTGCG	GTT CGATT	1380
CGTATTTCG	GTTGCGCG	GGTTAGTTA	CGCGGCGC	TTTTTTCG	CGCGCTTCG	1440
GTTTATTAT	AATCGCGTC	CGGGAGTGT	TTTCGTATCG	TTTCGGTGC	GTGGCGGG	1500
GATATAAGT	TAGTTAAGTT	TATAGTTC	GGTCGGGTG	GTTATGGAGA	GGCGTATTGT	1560
AGAGATATT	GGACCGTAT	TGTGATAGC	TTTATTTGA	TATACGTT	TTTTTATT	1620
TAATATTAC	GGGTCGTAAG	CGTTTATATT	TATATTATT	TGTAGGGATT	ACGTGGGAGA	1680
ATTGAAGGG	TTGATTTGT	TCGGGTGAT	TCGTAGTAGT	AATGATGCGA	GAATTCGTT	1740
TAATTCGTT	ATTGATTGGA	TTAGATTGTA	AATTTTTGG	CGTTGGGAT	GTTTAGTCG	1800
TGTTAGTAT	TAAGTGTAGG	TATTTAAATT	TTTGTGGAGT	AAGT GATTG	ATGAATGAAT	1860
GTATTGGAAT	TGTTTATTGA	GTATTTGAA	TAATTTGAT	TTTATGGTT	GGTATAAAGA	1920
ATTTTATTGT	AGTATGAAAT	AAAATGGTGT	TTGTGTGTT	TTTTGTTT	GGTTTAAGGT	1980
TTAAATAGG	TTAATGTATG	TTTTTTTTT	TTAACGTAG	AGGAGTTTA	TTTTGTTT	2040
GTGTTAGTT	TTATTATAGA	ATGCGTAATT	TTATTTTAAT	TTAAATGTT	ATATTTTTA	2100
GGGTAGGAAT	TGGTGTGTTAG	AATATATAGT	TTGAAGTATA	TTGTGTGTT	ATATAACGAT	2160
AGTTTTTG	GTTGGTAAG	ATTTTTTATT	TTTATTGTT	TTGTTATT	TTTTTATTAC	2220
GATTTTTTG	AATTTTTATT	TTTATTTTT	TTGTTTTTA	TTATTATTAA	AATTTTTTA	2280
TTGTTTTTG	TTTTTTTTT	CGTTAGTTT	TGGGGAGTT	TTGTTATGA	GTATTAGTT	2340
TGTTTTAGA	TGGTAAAGG	TTAGTGTAT	ATAGAAGATA	TATATTAGT	GATGGGTAAG	2400
ATGTTGTAAT	TATATTGGGT	TATGTTGAAA	TTGTTGAGTT	TTAATT	TATGTTGTAAG	2460
AGAAATTAAT	GTATTGTAAT	TTCGATGTT	GTGTTGAGGA	TTTTTTGTT	AAATTAAAGT	2520
TTTGTTTT	TGTTT					2536

ANGABEN ZU SEQ ID-NO:11:

SEQUENZCHARAKTERISTIKA:  
 LÄNGE: 504 Basen  
 ART: Nukleinsäure  
 STRANGFORM: Einzelstrang  
 TOPOLOGIE: linear

ART DES MOLEKÜLS: chemisch vorbehandelte Genom-DNA

SEQUENZBESCHREIBUNG: SEQ ID-NO:11:

AGTAGTAGTA	GTAGCGCGT	GAGTTCGGTT	ACGTAGAGCG	TTAGTTCGGT	GGCGTTTTGT	60
AGGTCCGCG	GCGT GAGTAG	CGGTAGCGC	AGTATCGAGG	CGTTAGTAG	GAAGGTGGCG	120
AGGTGAGAC	GCGTTTTAGA	GTTCGGCGG	GTTCGCGTCG	CGTAGTTCGT	TATGGCGTCG	180
GGCGCGGAT	TAGCGGGCGG	CGGCGTCGG	TGAGGTTAGTT	GTTCGTAGTT	GGCGGTGCGT	240
TGCGTATTGT	GTCGTCGATC	TCGGTTCGGG	AAGGGTAGTT	GGCGGTGCGT	GGCGGTGCGT	300
TGTTAGTTT	CGCGAGAATT	TCGGTTCGTT	TTGTTGTTGT	TCGGAATTG	GCGGGAGCGT	360
TTCGTCGTT	CGTTTTTTT	CGTTTTTCGG	GGATATTGCG	GTGTTGAGTT	AGATTTGTG	420
TGCGGCGGGG	GTCGGGAAGT	GGTGGGAGAA	GTCGTCGTC	GGCGTTGTT	AAATTAGTT	480
TTAAATTTAG	GAGGTTTTG	TTT				504

ANGABEN ZU SEQ ID-NO:12:

SEQUENZCHARAKTERISTIKA:  
 LÄNGE: 2036 Basen  
 ART: Nukleinsäure  
 STRANGFORM: Einzelstrang  
 TOPOLOGIE: linear

ART DES MOLEKÜLS: chemisch vorbehandelte Genom-DNA

SEQUENZBESCHREIBUNG: SEQ ID-NO:12:

AGTAGTAGTA	GTTTTAATTG	GAGTTAGCG	TATTCGAAGT	TCGTTACG	TTTTTTTTT	60
TAGAGATTT	TAAGTCGGAG	GAAAATAGCG	TAGTGTGCGT	TTTAGTTTA	TCGTTGTT	120
TTTCGGTTT	TAACGGITCG	GATAAGATGT	AGATGGAATT	ATCGTTTTG	TTAGGTTGA	180
GTTTTTATT	GTTTTTTAG	TAGTCGTCGT	CGTTTTAGGA	GTTCGCGGT	TCGGGCGCGT	240
CGTTGTCGT	GTTTTTCGGT	AGTATTTGGT	TTACGGGTAT	TATTAACGCG	GTAGAGGATA	300
GTTTTTTA	GGGGATTATT	TTAGTTAACG	GGATTATGTT	TTTTAGAAT	TTTCGTATT	360
ATGTTAATT	AGTTTTCGGA	GGTATTTTT	TTTCGTAGAT	CCGTTTGGCG	TAGATTTAGT	420
ATTATTAGTA	GTCGTCGTCG	TTTGCCTTCG	CGTCGTAGTC	GGTATAGTTA	CGCTAGTTAT	480
TATAGGCCTA	GTTTCGTAG	TAGCGTCGT	TATTGTTAG	TTTTAGTTAG	GCCTTTACG	540
CGTAGAGGAG	CGTCGTCGCG	GCGTACGGTT	ATTAGTTAT	TATGATTAGT	AAGTCGTTT	600
CGTTTCGGC	GGTTGTAGTC	GTTGTTGTCG	TAGTCGTCGT	TCGTTAGTT	TCGTTAGTT	660
GGAATACGTA	TTAAAGCGTG	AATGTAGTT	GGAGCGTATC	GTTTAATT	TGGGGCGGTT	720
TGTAGGCCGG	TCCGGGATTTT	CGTCGGCGG	TCGGTGTGGG	CCTGGGTGTG	GGTGTGGGG	780
TGTTTTTTTC	GTTAATT	ATTCGTCGT	TTAAAAAGTT	TTTTTTAGT	AACGTGATCG	840
CGTCGTTAA	GTTTTTCGCG	GCGGTTTTT	TTATTTTAA	GTTTTGGATG	GAGGATAACG	900
TTTTTCGGAT	CGATAATGGT	AATAATTGT	TGTTATT	GGTAATATT	TGTTGTATT	960
ATATATTTT	TTTTTTGTT	TCGTTTTT	TGTTATT	TTGATTTTA	ATTATTTATT	1020
CGTTTATATT	TAAGGAGG	AGTAATGTT	AGTTTTTTT	TTTTTTTCG	AAGTTTTAG	1080
TTTTTTAGGG	TTGTTTATT	ATTAGAGGAT	GAGGTTGGGA	GAATATTGTG	ATTATTGGAG	1140
GAATCGTATT	CGTTTTTTA	GGGTAGAAGA	AGTTTTTTT	TTTAGTTTA	TTTTTTTTT	1200
TATGGTGT	GTATTTTTA	TTTGATTT	TTTGATAAG	TAAAGTTGTA	AGTGTGTGGT	1260
AAGGTGTCGG	TATAGTTTA	GGATGAGTAG	GTTGAGATT	TTATTATT	GAGTAGTTAT	1320
ATATTTAGG	TTATGTAATT	TTGAGTTAG	GGTGTGTT	GTAAGCGGT	TATATTAAA	1380
TTATTTGAT	TTTATTTTA	AAGAGTTAA	TAATTTGA	GTGGTGTAT	TTAGACGGT	1440
TGCGTATGTT	TAGTTAATT	AATGTGATA	TATGTGTTA	TGTTTCGTTA	AGGTGTTAGA	1500
ATTATTAATA	ATAATTAGTA	TATTGTTT	GTTGGGAAA	TTATGAGTGT	GAGATTTAA	1560
TAAATATTTA	TTATTGTTA	AGGGAAAGGAG	GTGGAAGAGT	GGAAATT	AGGGTAATT	1620
TTTTATGTT	ATTAGAAGA	AGGATT	TTTTTTTTT	CGGAGGTAAG	AGATAATAGG	1680
ACGTGATT	TAGAGTTA	TTGTAATGGA	GTTTATTGTT	TAGTTAGTAT	TGAAATATT	1740
AAAGTTGTA	TGTTTTTAT	TAGTTGGTT	AAGAAAAT	AGATTAATT	TATTATTGA	1800
AATTGGTAA	AGTGAAGGTT	TTGTAATT	AATGAAAAT	GAGTTTTG	TTTGTTGTT	1860
AAATGTTAAT	AAAAAATT	GGATGTTT	TAGATAAAA	GGTTGAAATG	AAGAGTGT	1920
TTATTTGTT	TTTGTTTT	TGATTTAGTT	TATAGTTATA	GTAATTGGTA	GTGTTTGGG	1980
TATTTGGGT	TGGAAAGGCG	TAGTGCGGGT	GTGTGTTAT	AGGCGGTTT	TGTTT	2036

ANGABEN ZU SEQ ID-NO:13:

## SEQUENZCHARAKTERISTIKA:

LÄNGE: 452 Basen

ART: Nukleinsäure

STRANGFORM: Einzelstrang

TOPOLOGIE: linear

ART DES MOLEKÜLS: chemisch vorbehandelte Genom-DNA

SEQUENZBESCHREIBUNG: SEQ ID-NO:13:

AGTAGTAGTA	GTAATTTTT	TGTATTTGTA	GTAGAAAATT	TATTTAGAG	ATTGGGTATG	60
TGTGTTAAGG	AGATGGTGT	TTTTTTGGA	TATTTAGAG	ATTTTTTTT	TTGTTATGAT	120
TTTTATTGCG	GGGATATT	TATAATGTA	TTTTTTGGA	TAGTTTTGT	TTATATTAAAG	180
ATTGTTATT	TTTAGTTGT	TTGTTGTT	TTAGATTATG	ATATTATT	AGTTGTTAG	240
TTTAGGAAAG	ATTTTGGGT	GTATAAATT	GTTAGAATA	GTTGGTAGTA	ATTGTTAGAT	300
CGTTGTTGT	GTCGTGAGGG	TATGGATGTG	GTATATT	ATGGTTATAT	TTATATT	360
GGGGGACGAG	ATTTATTAT	TGGAGTTAAG	TTGAAGGAAG	TGGAATGT	TAGTGT	420
AGAAATTAGT	GGGTATTG	GGTTTTGTT	TT			452

ANGABEN ZU SEQ ID-NO:14:

## SEQUENZCHARAKTERISTIKA:

LÄNGE: 513 Basen

ART: Nukleinsäure

STRANGFORM: Einzelstrang  
TOPOLOGIE: linear

ART DES MOLEKÜLS: chemisch vorbehandelte Genom-DNA

SEQUENZBESCHREIBUNG: SEQ ID-NO:14:

AGTAGTAGTA GTAGCGGAGA CGGAGGGCGC GCGGGAGGCC GCGGCGGCCA TTTAGCGCG	60
GGGCAGCGG CGGGGTTTAT GGTCGGATTG GCGAGTTGG TTTAATAGTG GGGTTAGGGG	120
TTTCGGGCGC GGGGGGGATC GGAGGGAGCG AGGTCTTGT CGGACGGGGC GGGCGTCGG	180
GAGGGGGCGGG GTCGCGGTG GACGGGGCGG GGTATTAGGA GGAACGGAGT GGGCGTTGG	240
CGGTTTTCGC GTTTAGATTG GGTCGCGGGC GTTTTTGGT GGTCGCGGAG AGTTAGGTG	300
TTCGCGGTG AGGGAGTTGG AGAGGGTAA AATCGGGTC GTAGCGGAA GGCAGAAAGT	360
TAGAGAGAGG AGGTGTTTCG GTAGCGGACG CGTTAGCGGG GTAGATAGGA AAATAGTTGG	420
AAAGGGCGAT TTGGGGGAGT AGAGATACGA TTAGAGTTTG AGAAGGGTAG GAGTAAGGGA	480
TGTTTAAGGG GTTTGGTGGT AGGTTTTGT TTT	513

ANGABEN ZU SEQ ID-NO:15:

SEQUENZCHARAKTERISTIKA:  
LÄNGE: 980 Basen  
ART: Nukleinsäure  
STRANGFORM: Einzelstrang  
TOPOLOGIE: linear

ART DES MOLEKÜLS: chemisch vorbehandelte Genom-DNA

SEQUENZBESCHREIBUNG: SEQ ID-NO:15:

AGTAGTAGTA GTATTTTCGG TTATAACGGAT TTTTCGAGTT TCGTTCGTCG GTCGCGGATT	60
CGGTTTTTTT TTTTCGGTC GTTAGGGGC GGGGTCGGAT TATAGGATTG GAGTTGGCG	120
GAGATTACG TTCGGAGCGG TTGTGAATTG GTAGGGCGGT GGCGCGGTTT TTGTCGCGT	180
TTTCGGGTAT TTAGTTTTT AACGGGGTTT CGGAGTCGAA GATAGTTTA GGGTTAGGG	240
AGCGGGGCG GTTTTGGGC GGCCTTAGAT TGCGGTGAGT TGGTCGGGT GGGTTATTAA	300
TTTAATGTAG TTTAGGGCGG CGGTACCGAGA TAGAATAACG GCGAATAGGA GTAGGGAAAG	360
CGTTTCGAT AGGTTAGGTT TAGGGATTG CGGGGAGAGG GCGAGGTAA TATTCGGTAT	420
GGGTTTTGA TTGGTTTTG GGATTCGTT CGTTACGTT TATAGGTGGG TTCTGATTTT	480
TTTTGCGTT TCGTTTCGT TTTAATAGTT TATAGTTGTT GTTACTTTAT TCCTACGTT	540
CGAATTTCGT TCGAATTTCGT TATTGGTTGT TTTTACGGG TTTGTGTTGA TTGGTTGTT	600
GAAGATTTCGT TTTTTGTCG TGGGTTAGT TTCTAAATG CGTAGTTAAG CGGGTGGTAA	660
GGGGCGGGTG GAGCGCGGGG CGCGACGGCG GAGGGGGCGG TGGGTAGTCG GACGTATTT	720
GGTAGGGAGT AGTAGGTGGC GGCCTGTAT GGGGTTGGT TTTATTAGCG GGTATTGGTT	780
TATAGTTACG GTCCGGGGGGT TATTTAGTTG GAGAGAGAAG GGATAGGTGA TTCTGATCGGA	840
GTTTAGTTA GTTTTAGCG GTGGGGCGAG AGATAGCGAG GGAATCGAG GTTGGGGAGG	900
TTATTTAGGG AGATTTCGGA GGGAAATTGG TGAGGTTGA ACGGAGGGAG ATTTGGGGTT	960
GAATAAAGGG TTTTGTTTT	980

ANGABEN ZU SEQ ID-NO:16:

SEQUENZCHARAKTERISTIKA:  
LÄNGE: 223 Basen  
ART: Nukleinsäure  
STRANGFORM: Einzelstrang  
TOPOLOGIE: linear

ART DES MOLEKÜLS: chemisch vorbehandelte Genom-DNA

SEQUENZBESCHREIBUNG: SEQ ID-NO:16:

AGTAGTAGTA GTAATATTAT TTTTAAAAA AAATATTAGC GTAGTTTTA ATAGAGTTT	60
ATTTTTAAA GTGAAGTTAA AAATAGGTTT TATGTTAAA GTTTTTTC GTAGTAGGCG	120
ATACCGTAG ATGGAGAAA TATTTTATT GTGGGAAAG GGAGGCGTGT TAACGGGGAC	180

GAAGGGATATT TTATTATTTG GTAAAGTTAT TGGTTTTGT TTT

223

ANGABEN ZU SEQ ID-NO:17:

SEQUENZCHARAKTERISTIKA:  
 LÄNGE: 1145 Basen  
 ART: Nukleinsäure  
 STRANGFORM: Einzelstrang  
 TOPOLOGIE: linear

ART DES MOLEKÜLS: chemisch vorbehandelte Genom-DNA

SEQUENZBESCHREIBUNG: SEQ ID-NO:17:

AGTAGTAGTA	GTTTGGGTT	TCGTTAGTC	GTAAGATGAG	TTAGTTGGGT	TTGGTGGGGG	60
TTGTATTTAG	GTACGTTT	TTAGATTTCG	TTTTTTTTA	TTAGATTACG	TTTATCGTT	120
TTTAGTCGT	TTTGTTTA	GGTTTTTTT	TGTTTTTTTC	GTTTGAATT	TTGTTTCGAG	180
TTTCGTTAT	TTTTTCGCT	TTTTTTAGT	TTGTTTTTC	GGATTTGGT	CGAGTATTT	240
AAGTTTTT	TTGGCGATT	AGGTTTTT	ATTATTTAGG	TTTCGTTTT	GTTGTTTCGT	300
TTGGGGCGGT	TTGAGTTTT	ATCGTCGAGG	AAGAGGAAGG	CGTTCGTGTT	TTGGTTTCGT	360
AGCGATTCGT	TTTCGTCGTA	TCGAATTCTG	AGTTTATTAA	TTATGTTTAT	TACGTCGATA	420
TTTATCGTTA	GTATTATGTA	CGTTATAGTT	ATCGTTAGCG	TGGTTGGGGA	GAGCAGGCCT	480
ATTAGTTTC	GCAGGGCGTCG	TTTTAGTTAT	TACGGGTTCG	GTCGTAATCG	GAAGTTTTT	540
CGTTTTTGT	TTTAATTTCG	GTGTGTGTT	AGGAGTCGAG	TTTTATCGAT	CGGGAGTTAG	600
GCCTTGGAT	TGAGATTG	AGGTTATATT	AGGATTGCG	GTTATTGTCG	TTATTAGGAC	660
GGGGAGGGGG	TGGGGGATAA	GTTTATAAGT	CGAGGGCGG	GGTTCTGGC	GGTTTCGGTC	720
GATAAGGGAT	AGTTGTAATT	GAGTTCGGGG	TGGAAGGGGT	GGGCGTTTCG	TTATGGGTTT	780
AGTTGTGTG	TTCGAGTTTT	GGGGGTTGA	AATTTTTT	CATATTG	CGGGCGGAGT	840
TTTATTTG	TAGGGGATGT	TGGAGTAGAT	GTCGTTGTT	TATTTTGT	ATAGTTATT	900
GTGGTTTTT	TGTGGCGGG	TTAGTAGTT	GGTGGTCGTG	GAGACTATTA	TGAGGACGTT	960
GGTTACGAGG	TTGAGTAGAA	TGTTTTTATT	TTGGAGGTTT	CGTTTTATT	TTATGTTATT	1020
GAGGTTGAG	TCGGGGGTTA	TAAGGGTAGG	ATGGGTTAG	GTTTTTTTA	TTTCGAGGTT	1080
TTTAGTTGAG	GAGTCGTTT	TTTTTATATT	TTTTTTG	TATTTTGT	TGAGGTTTT	1140
						1145

ANGABEN ZU SEQ ID-NO:18:

SEQUENZCHARAKTERISTIKA:  
 LÄNGE: 633 Basen  
 ART: Nukleinsäure  
 STRANGFORM: Einzelstrang  
 TOPOLOGIE: linear

ART DES MOLEKÜLS: chemisch vorbehandelte Genom-DNA

SEQUENZBESCHREIBUNG: SEQ ID-NO:18:

AGTAGTAGTA	GTATAGTTCG	GCGTTGGTTA	GCGGTTAGA	GTCGGGGTAG	GGAGGGTTTT	60
CGTCGGACG	GGGGTTAAGA	GGTCGTTTT	AGTTCGGGT	AGTCGCGTA	TTCGAGTTG	120
GAGTTCGGAG	TTGGGAGGGG	TGGGGAGGTT	TCGCGTTAG	GGTCGGATCG	GAGGGAGAGG	180
GAGAGCCTG	GTTTTTTTT	CGGTTTTCTG	CGTTAGTCGA	TCGGGGCGTT	GGCGGGCGT	240
CGCGGGAGTC	GTAGTTTTT	TCGGGGGGCG	GATTGGTT	CGGCGGCGGT	ATTTAGTTT	300
TGTTTCGATA	GGTGCCTCGTC	GCGCGAAGGA	ATGTAGTCGG	TTTGATTTAT	TAGCGTTTT	360
TTTTTATTG	CGCGTTCGTT	TTATAAAGCG	TTGTTCGTT	CGTTTTTATT	TTTTTAATT	420
TCGCGTCGT	TTTCGAGATAG	TTTTGTTCG	TTCGCGCGTT	GTAGTTTAT	TTTTTAGCGG	480
TAGTTAGGC	GCGGAGGGAG	CGAGTCGTT	TGAGGTAGG	TTAGGACGG	GCGTATAGTA	540
GTAGTCGAGG	TTGGTCGGGA	GAGGGTAGT	GTTCGTTTA	GTTGGGTAGT	CGCGTAGGGG	600
AGTTGGTCGG	GTAGGGTCGG	TGGTTTTGT	TTT			633

ANGABEN ZU SEQ ID-NO:19:

SEQUENZCHARAKTERISTIKA:

LÄNGE: 12 Basen  
ART: Nukleinsäure  
STRANGFORM: Einzelstrang  
TOPOLOGIE: linear

ART DES MOLEKÜLS: chemisch vorbehandelte Genom-DNA

SEQUENZBESCHREIBUNG: SEQ ID-NO:19:

AGTAGTAGTA GT

12

ANGABEN ZU SEQ ID-NO:20:

SEQUENZCHARAKTERISTIKA:  
LÄNGE: 12 Basen  
ART: Nukleinsäure  
STRANGFORM: Einzelstrang  
TOPOLOGIE: linear

ART DES MOLEKÜLS: chemisch vorbehandelte Genom-DNA

SEQUENZBESCHREIBUNG: SEQ ID-NO:20:

ACAAAAAACTA AA

12

ANGABEN ZU SEQ ID-NO:21:

SEQUENZCHARAKTERISTIKA:  
LÄNGE: 74 Basen  
ART: Nukleinsäure  
STRANGFORM: Einzelstrang  
TOPOLOGIE: linear

ART DES MOLEKÜLS: chemisch vorbehandelte Genom-DNA

SEQUENZBESCHREIBUNG: SEQ ID-NO:21:

AGTAGTAGTA GTTTCGGTAG TTTAGTTAT GGCGGCGGTG GCGGCGGTAG TAGGTTGAG  
TTTTAGTTT TGTT

60

74

ANGABEN ZU SEQ ID-NO:22:

SEQUENZCHARAKTERISTIKA:  
LÄNGE: 103 Basen  
ART: Nukleinsäure  
STRANGFORM: Einzelstrang  
TOPOLOGIE: linear

ART DES MOLEKÜLS: chemisch vorbehandelte Genom-DNA

SEQUENZBESCHREIBUNG: SEQ ID-NO:22:

AGTAGTAGTA GTAGCGGTAG CGGTAATAGG GCGGTTGAGA ATTCCGGCGGC GGCGTTTTT  
TTCGTTTTT TTTTTTCGT TTCGTCGATT TTAGTTTT GTT

60

103

ANGABEN ZU SEQ ID-NO:23:

SEQUENZCHARAKTERISTIKA:  
LÄNGE: 559 Basen  
ART: Nukleinsäure  
STRANGFORM: Einzelstrang  
TOPOLOGIE: linear

ART DES MOLEKÜLS: chemisch vorbehandelte Genom-DNA

SEQUENZBESCHREIBUNG: SEQ ID-NO:23:

AGTAGTAGTA GTAAAGAAGGA	AAAAAAATAAA	TTTTTTAAGT	ATTTAGTGTA	GTTTTTTTG	60
GGTTTTTAG TTTTATTTCG	GTATAGTTAG	GGTTTTAAT	TTATTTTGT	TTGGGTTTA	120
GGGAAAATAA AAAGTTAAGT	TTATTCTAT	TTAGTTAAAAA	GTTGAAAATA	AGTGGTCGGA	180
TTAATTTTG GTTTAAATA	AATTATTGA	TGGTGTAGA	AGGTAATAAA	TTTATTTT	240
TTTTTAGAT AAGGGTTAT	TTTACGAGAT	TATTAAGTAT	TGTTTCGTGA	GATTATTGAA	300
ATAATAGTAT TTTTTTTT	ATCGAGGATT	ATAGTATGTT	TTGATTAA	AACGTTTAA	360
GTTTTGAAG AGATATTGG	GAGTTTTAG	AAATTAAGG	AAGGATTGGA	AATGTTAGGG	420
GGTGAGGGC GTTAAATAAT	TTTGATTTT	ATGGTTGTGA	TTGATGTTGT	TATTAGATTA	480
TTTTTATATT TAATAATATT	GAATTATTA	AACGGATTAG	TATAAATTAA	ATTATGATAG	540
TATATTTTA GTTTTGTT					559

ANGABEN ZU SEQ ID-NO:24:

SEQUENZCHARAKTERISTIKA:

LÄNGE: 1695 Basen

ART: Nukleinsäure

STRANGFORM: Einzelstrang

TOPOLOGIE: linear

ART DES MOLEKÜLS: chemisch vorbehandelte Genom-DNA

SEQUENZBESCHREIBUNG: SEQ ID-NO:24:

AGTAGTAGTA GTTTGGAGA	GTAAGTGGGT	TTTATGTAAA	GTGTTATTG	TTATATATTG	60
TTATTTTAN AGGATTTAG	AATATTTTCG	AATCGAAAA	TAGAGTGAGT	GTGGAGGGGA	120
GGGGGGTGT	TACGGGAGG	AGGGTCGCGA	GAAGGAGAGT	GCGTCGGGTT	180
GTTTTCGG	GGGTGATGT	CGGGCGGGAT	CGGCCGCGCG	GGGGTGGGGC	240
CGGCCTCGA	GTAGTAGAGG	GGCGTTTCGT	AGAGGTCGCGG	GGGGGCGCG	300
CGTTTGTAA	GTAGTCGTT	CGTCGTTTT	TTCGGGTCGT	TGCTAACATCGT	360
GTCGGTTCGG	GTTTTGAA	TTTTTACGG	GGGATAGTCG	ATCGGGGGCG	420
TTTCGTATT	CGATTCCGAC	GGCGGTAGTA	GGGGGAGGGA	TGGTGTATCG	480
GGTTTTATTA	TTTGTTCGG	CGGGTCGGCG	GTCGTTGGGT	TCGTTGGTC	540
TGCGTTTAGG	TTGGTTCGTA	GGAAACGGCG	GCGGCGGTTT	GGAGAAAGTGT	600
TNNTTGNAN	TNTTTTTTA	TAGGGGAGGG	GTAGGGCGTT	CGCGGGTTTC	660
GGCGATTGGG	TACCGCAGGG	AGCGGGTAGG	GTAGGGGGAA	ATAAATTAAG	720
AGAGTCGGAG	AGCCCGGGAG	GAGTAGCGGC	GAGAGGTAGG	AGAGGTAGAG	780
GAGGGAGAGA	GGGGCGAAG	ACGGTCGGGA	AATTGGGTT	TGTAGTTCG	840
GAGTCGTGAC	GGATTTATTG	TAGCGGTGCG	GTTGATTGTT	TACGAAAAAA	900
GGGGAGGGTA	GAGAGGGAGG	GGAAACGGGA	GTGTTGTAGC	GTCTGGGCGG	960
TCGCGACCGT	CGTATATTG	GGAAATTGCTN	GTTTCGTTGC	GGCGGGTTT	1020
GGCGAGTTT	GGTTGGGGAA	NTTGAGCGC	GCGGATTACG	TTCGTTTTT	1080
TTTTTGCCTG	TAGGGTGGGG	GGAGGTTAGG	GGGGGGGGCG	CGCACGTCGG	1140
TGGCGGAGTT	TTTCGGTTA	TGTGGTTGGA	GGCGGGCGGG	TGACGTCGCG	1200
GGGAAGGAGG	AAATGCGAGG	GTTCGTTGCG	GGGAGGAAGC	GCGAGGGGTG	1260
AGCGAAAGGG	GGATAGGGGT	TTGGAGTGG	AAGTTTGCG	GGAGAGGAGG	1320
TAGTTTATTAGG	GTGTTAGAAT	TATTCGGATG	TCGTATTAAA	AAATAGGAA	1380
TTGCGAGAGG	TAGTAAAGG	TATTTGTGTA	TTCGTGTATC	AATTAAAGTT	1440
TTGTACGGTT	TATAGGATT	GTTGTTATT	CGGTGTAGTA	TTCGTAGGTT	1500
ATGAATTGGT	AGAGTTAGGA	ATTAGGGTGT	AAAGATAGAG	AAGCGAGTTA	1560
TAGTCGAGG	CGTGGGGAGA	ATTGGGTAA	ACGAGGAGAA	GGGATATTTT	1620
AAAGATTTTT	TATTTAGTTT	TGTATTTTA	AAATCGTTT	AGATTTTTT	1680
TTATTAGTTT	TTGTT				1695

ANGABEN ZU SEQ ID-NO:25:

SEQUENZCHARAKTERISTIKA:

LÄNGE: 722 Basen

ART: Nukleinsäure  
 STRANGFORM: Einzelstrang  
 TOPOLOGIE: linear

ART DES MOLEKÜLS: chemisch vorbehandelte Genom-DNA

SEQUENZBESCHREIBUNG: SEQ ID-NO:25:

AGTAGTAGTA	GTTCGTTAC	GAATACGATT	AGTTTATATT	GGTTGTTAG	TATTCGGTT	60
AGTTTTGTG	GGACGTTGGC	GCGTATATCG	TTTCGTCGGG	GGGTCGGGGA	TTTCGGGGAT	120
TTCGGGGTCG	AGGACGAGGG	AGGCAGTAG	GTCGTTGGTT	TCGACGATT	TGCGATCGAG	180
GCGTTTGTAG	TTGTCGGGG	CGAGGCGGGC	GATTTAGGT	TTTCGAGTAT	TTTCGAGATA	240
TTCGCGTTAA	GTTGGTTTC	GGGGGTGGTT	GTCGTCGGGT	TTCGTTGCG	TTTTTGGTTT	300
TTGTTTTGT	TTCGACGTTT	TTGTTAATT	GTATTTTTT	TTTAATT	TTTCGATTT	360
TTATTTACG	TTTTGTTAA	TTGTTTTTT	TTTTGTTAT	TTTCGACGT	TCGTTTTTT	420
TTTTTTGTT	TTTCGTTTT	TCGTTAACGGT	ATTATTTGT	TTATTTATTT	AGCGTTTTAT	480
TTTGTGATT	TGGGATTTA	CGAGTTTTT	TGTCGTTGT	TTTTTATTGG	GTAACGTTCG	540
GGGTAGTTAT	TTGTTTTTT	CGGGATTAC	GCGGATAGTT	TTTCGTTTT	GATTTTGCGG	600
ATATTGGTTA	GTTCGTCGC	GATATTAGCG	CGTNNTTTCG	TATTTTCGTT	CGCGCGTTCG	660
GTTCGTTTC	GTCGTTGTA	GTTTTATTT	TTGAGTCGAC	GTTCGTTTAT	TTAGTTTTTG	720
TT						722

ANGABEN ZU SEQ ID-NO:26:

SEQUENZCHARAKTERISTIKA:

LÄNGE: 517 Basen  
 ART: Nukleinsäure  
 STRANGFORM: Einzelstrang  
 TOPOLOGIE: linear

ART DES MOLEKÜLS: chemisch vorbehandelte Genom-DNA

SEQUENZBESCHREIBUNG: SEQ ID-NO:26:

AGTAGTAGTA	GTCGAATTC	GCGCGCTAG	CGCGTATGG	TTAGGTTAG	CGAACGTAGT	60
TTAGCGAGT	GGCGCGTTAG	GCGTATTACG	TAGAGTACGC	GTAGCGTAGC	TAGTAGTCGT	120
AGTATTAGTT	TCGCGCGTTT	TAGGAGTTG	GTTCGTTCG	GGTTTGTCTG	TAGTTTTAGT	180
AGTAGCGATA	CGTAGAACGG	TAGGAGCGTT	AGGATGTTAA	TGATGTTGAG	TGGCGCGCGT	240
AGGAAGGCCT	ATTGTTTTC	GGTTTGTAGG	GAGCGTAGTA	GGAATTGCAA	GGAGAATTAG	300
GTTACGTATA	CGGTTTTAG	TACGAATAGG	TTGCGGTATT	TGGGGGAGTA	TTCGTTTGC	360
GGGGAGAGGG	GATACGGGGT	TGGGGCGTA	GGTTTTTGG	AGGGTCGGGG	GCGTTTGTT	420
TTTTTTTTT	TAGGATTAGA	GGC GTTTTT	AGGTTATTT	TTTCGTTTA	ATTTACGGAA	480
GGGGGTGTAG	GGGATATTGA	TAGGTTAGT	TTTGTT			517

ANGABEN ZU SEQ ID-NO:27:

SEQUENZCHARAKTERISTIKA:

LÄNGE: 1078 Basen  
 ART: Nukleinsäure  
 STRANGFORM: Einzelstrang  
 TOPOLOGIE: linear

ART DES MOLEKÜLS: chemisch vorbehandelte Genom-DNA

SEQUENZBESCHREIBUNG: SEQ ID-NO:27:

AGTAGTAGTA	GTATTAGTAA	TAGTAGTACG	AAAAGAAAAA	TTGTAATTAA	AACGGTTTTT	60
AGGGTTTAAG	TAGGTCGAC	GAAGATTCTG	TTATTCGGTC	GTTAGAAAAT	TGGGAGTTTC	120
GTTCGTTTTT	TCGGTATTGA	AACGCGATCG	GTTTGTTTG	GTATCGTATT	TTTTTTTGA	180
TTTTATTTAT	ATTACGACGA	CGGACGTTCG	AGAACGTTGT	ATCCGTTTC	GTAGGAAGTG	240
TTTTTTGGG	CGGAAGTTT	TGAGCGTGT	ATAGCGGAAG	TGTTTTTTT	TTCGTGTGTT	300
TTGGTTTCGG	TCGTAGAACG	GAGATGGTGA	GTTGTGATTG	TGGTGTGTT	GAATCGCGTT	360

TTATTTTCGT	TTTTTGTGTT	TTTTGTTTG	TTGTGTTGG	GGGGTTGGTA	AGATTTCGGA	420
TAAGGGAAAT	TGGTGGGTTG	GAAAGAGGTA	TGCCGTGGTT	TTAAGAGTT	AGAAGAACATGA	480
TTGTTAATTG	GTGTTTGGGG	GATTTATTC	GTCGTAATTG	TGGTGTAGA	GTCGTATTGT	540
GTTCGGTTGTT	TCGTTTAAT	TTTGGAGAT	TTTACCGGT	TTAGTGTGTT	GTTGGGAGTC	600
GAGGGAAAGGA	GTTGGGAAT	GTTGGTTT	GTGTAATAAT	GAAATAATT	ATTGGTGTAG	660
TTTTGGTC	GGAGTTGTA	AAGATAAGGT	GTATTTAGA	ATATTGTAAT	TTTGCAGGAG	720
GGTTTAGGT	AACGTAAAT	GCGGGTAGTG	GTTTTGATT	TGGTATTGCGT	GGAAAGAGGT	780
TTTTTCGTA	GTTCGATT	TTACGATGT	TTTAAATT	TATTAGTAAT	TGGTTTTTCG	840
GAGAATCGA	GTAATTAG	AAAGTTAG	GTTTAGAAT	TATTTATT	TTAATGTGT	900
AGACGAAGGG	AACTTATCG	TTGGAAAGC	GTCGTAATAA	GACGTATAACG	TTGTGTCGTC	960
GTTGTGGTT	TAAGGTTAT	TATTTTAA	AGTCGATTG	TGGTAAATGT	GGTTATTTG	1020
TTAACCGTAA	GAGAAAGTGT	AAGTAATATT	TTTAGGTTA	ATTGTGTAG	TTTTGTT	1078

ANGABEN ZU SEQ ID-NO:28:

## SEQUENZCHARAKTERISTIKA:

LÄNGE: 2949 Basen

ART: Nukleinsäure

STRANGFORM: Einzelstrang

TOPOLOGIE: linear

ART DES MOLEKÜLS: chemisch vorbehandelte Genom-DNA

SEQUENZBESCHREIBUNG: SEQ ID-NO:28:

AGTAGTAGTA	GTCGTAGGAG	TAGCGTTTG	GGGAGGGGGG	TTCGTTTTC	GGGGTGCCTT	60
TCGTGGTTAT	TTTTTTTAAN	GGGGGTCGTG	GTCGTTTTT	AGTTTCGGCG	CGTTGGGTTTC	120
GTTCAGCCG	GGTCGTCGG	GCGGGGCGGG	GATTGTTATT	CGGGGCGTT	TTTCGGGACG	180
CGCGGGTTT	CGTTTTTTT	TTTGTGTTCG	GGCGAGTTCG	GATTTGTTC	GGTCGCGGGG	240
TTTCGTTAT	TTGTTAAAGTG	AAGGTTTAC	GGGAGATAAT	AAAGAAAGAA	GTTGTTTTT	300
TTTTTTTAG	TTAAATTACG	GATGATTAGT	AATCGGTTCG	TTTTTTTTT	TGTCGTTTT	360
TTTTTTTTT	TTTTTTTTT	TTTAGGAAT	ATATTTATT	CGTTTTTTT	TAGAGTAGTA	420
AAATTTTTTC	GTTCGTTAG	GCGTGGTGT	GTTGTTGTT	CGCGAGTGC	GAGAGGGGAG	480
GTTGGTTTT	GTTCGTTTTT	TTTGGTTT	TCGTTTAAG	TTTGTGTTT	TTTGTAGAAA	540
ATTTTTGCG	ATGGTGTATT	TTTTGAACT	TGTTATTGTC	GGAGTTTTG	GTTGTTTTG	600
AAATAAGAGT	TTTTTTTAT	TTTTTCGTA	ATAGTTATAG	TTTGATGCGA	TATTTTTTT	660
TTTGGGTAGA	AATTAGGAA	AAATATAAT	AAGGAAAAAA	TTGAATAGGA	AAAAAAATGA	720
TGTCGAATCG	TTTGTATT	TTGTTGTTGA	CGGTTGGGGT	TGGTGTGTTA	TTTTTTTTT	780
AGTAGTTGTT	TGGAGAAGGA	GGAGGAAGAA	GAGTGTATTG	GTGGAGGTGG	AGGTGTGTGT	840
GTGTATATAG	GGGATCGATT	TTTGATATAT	AATAATTAA	GTGGGTACGT	TGTACGGTTC	900
GTAGAGTGT	TCGACGGCGG	GCGTGGCGT	TGGTGTGTT	GGAGTAGAGT	TGTAGTAGGA	960
GTTCGTTTG	CGTTTCGTT	CGCGTTTTT	TTTTTTTTT	TATTTTTTTT	TTTTTTTTT	1020
TTTTTTTTT	TTTTTTTCG	GGCGTTCGTT	TTCGTTTCGA	TTTGATAGTT	CGTTTTTGT	1080
TTTTTTTTT	TAGAGCGGGG	ATTGTCGAAT	TTTATTTTT	CGTCGCGAGC	CGCTATTAT	1140
TATTATATT	TTTAAATAT	AAGGAAGTCG	AGTATTAGGG	TTTCGTTGA	TTGGTTATCG	1200
TTGGGTGATT	GGTAGGTTCG	GAATGATTGG	TTTAGGACGC	GAGGTTTTT	TTTCGGTCGT	1260
TTCGATTGGT	TGTTTTTAA	TTAGGTAGA	GGCTCGTAGA	AGTTGAAAT	TGTCGTTTT	1320
TTAGTTTTT	TTTTTTTAT	TATTTTTTT	TTTCGTTT	TTTCGTCG	TTTTTTTCGT	1380
TATATTTAA	TTTTTAGTGT	TCGGTTAGA	CGTTGGCGT	TTTCGGCGG	TTTGGCGTT	1440
CGTAATAGGT	TTGGCGGGGG	GGAAGAAAGG	GGAGATAAAA	GGGAGGGAGG	GACGAGAGGG	1500
GGGAAGAGAA	TTAGAAGGAA	AACGAAGGGG	GAAATATGAA	AAATAGTAAT	TTGTTTATT	1560
TATTTAGTAC	GCGTCGGGTT	GTTTATTTT	TTATTTTCGT	TTCGTTAGAG	ATTGTAAAG	1620
CGCCCGTAC	GGATGTATTA	TTAAGTTAA	TATTTATAGA	AGATGTGTT	TTTAGTGGTT	1680
TTAAGTTT	TTTTTTTTT	TATTAGTCGG	TTTTTTTTT	TGATCGGCGG	GGGTGGGAAG	1740
CGGGGGTCG	TATCGTTAAG	AGTTTGGTC	GTGGTTACGT	TTTTGGAAA	TCGTTATCG	1800
TAATTTTCG	TAGTCGTTGC	GGCTCGGCGG	TAGTTTTT	TGTTAAGTT	TCGGATGTT	1860
ATTATTGGG	AAATTGTAG	TTAGGANNT	TCGGTGTGTCG	CGCGTGCAG	ACGGATCGTT	1920
TCGTGTCGT	GTTTTTGTT	TTTTTATT	TTTTTTTTT	TTTTTCGTT	TTTTTTTTT	1980
TTTTTTTTT	TTTTTTTTG	GGATTGCGTGT	AAGTAGCGTG	CCTGTTTTG	TGAGTTTGG	2040.
GGGTGTCGGT	TAGCGGAGTC	GTGCGATGTA	TTGAAAAGT	AATTAGTTG	ATTTTTTTT	2100
ATAGAGTTAT	CGTAAGTGT	TCGATTAA	ATGGTTTAG	AAATTTTTGT	GGTATTAGGA	2160
TTCGGTCGAA	AGAACGGGGA	TCGGTTATT	CGCTTTTTT	TTGTTTATC	GTTCGTT	2220

GTGTGTGTGT	GTGTGTGTGT	GTGTGTGTGT	GTGTGATTAG	CGGTGGGGGG	GTTTGTTCGT	2280
TTTTTATTTT	TAGGTATGGT	GGTGTTTTTT	TTTTTTTTTT	TTTGTTTTA	GATGGATTG	2340
CGTGGTGGAT	GGGGTTGGCG	GCGATAAAATG	TTTTTTTAGT	TTAATTAG	TTGAAAGAGT	2400
TAAGGGGAT	GGGAGGGGGA	GTGTATTCGG	TAGGCAGAG	AAGCAGGGGT	GGGGTGGGGT	2460
GGGGTGGGG	GACGCGGTTA	AGCGGAAACG	TTGTATAGAG	GAATTTAGC	GAATTAAGAA	2520
AAAAGAGAAA	GTGGTAACGA	AAATAACGGT	AAATTGAGTT	TTTTCGGGG	ATTTTAATG	2580
AATTAATTAA	ATTCGGATAT	TTAATAATA	TATGGTTTT	AATGAGCGTG	CCTGTGCGTG	2640
TATTCGTAT	TTTAGTTGC	GGGTGCGTTC	GTTCGCGTT	GTCTGTTTA	GTTAGAGTTT	2700
GTATTTGGTA	GTTCGAGTT	CGAGCGGTAG	TTTAGGACGT	AGTCGAGGAG	CGTCGTCGGT	2760
GCGTTTCGAT	TAAAATGTGA	ACGGGTTGT	TTTTTTTTT	CGTTGTATCG	TGTTTTGTC	2820
GAGCGTAGTT	AAGTGAATT	AGTAAAGTA	GGAGTTTTT	TGTTTAGTCG	TAGTTAGGTA	2880
GTTCGCGT	GATTTAAGA	TTAATATTAG	ACCGTAGAA	TTTATGTAGG	TTTTGTTTA	2940
						2949

ANGABEN ZU SEQ ID-NO:29:

SEQUENZCHARAKTERISTIKA:  
 LÄNGE: 117 Basen  
 ART: Nukleinsäure  
 STRANGFORM: Einzelstrang  
 TOPOLOGIE: linear

ART DES MOLEKÜLS: chemisch vorbehandelte Genom-DNA

SEQUENZBESCHREIBUNG: SEQ ID-NO:29:

AGTAGTAGTA	GTTTATTTG	GGTTGATATA	AAATTGAAAA	ATTTATTGT	TTTTTATTT	60
TTTTGGGTTG	GATTCGGTGT	ATCGGTTGAT	ATATTTTTT	GGTTATTAGT	TTTGTT	117

ANGABEN ZU SEQ ID-NO:30:

SEQUENZCHARAKTERISTIKA:  
 LÄNGE: 639 Basen  
 ART: Nukleinsäure  
 STRANGFORM: Einzelstrang  
 TOPOLOGIE: linear

ART DES MOLEKÜLS: chemisch vorbehandelte Genom-DNA

SEQUENZBESCHREIBUNG: SEQ ID-NO:30:

AGTAGTAGTA	GTATTATGGT	TATTCGACG	GTCGCCTCG	TTATTTTTT	GCGGCGGTT	60
AGTCGCGACG	TTGTTAGGCT	TAGTAAATT	TTTTTATTT	TGGGCGTAAT	GGTTGTCGGG	120
GTCGGGGTT	TTCGCGGGTT	AGGAGGGCGT	GATCGTTT	GGGTTGGGT	TTGTATTTT	180
TCGACGGTTG	TTTTCGTTT	TTTTTATTT	GTTCGCGG	GTTCGTTAGTC	CGGCGGTGTC	240
GTTCGTTAAG	TCGTTCGTTA	AGGGGAGGTT	TTTCGTGGGT	TACGTTAGGG	GTAGTTTCG	300
ACGGTTTACA	GTAGTGGTT	TTTTAGTATT	TTTCGTTA	GTTCGTTAACGA	TTTTGGGTAT	360
TTGAGATTG	CGGTTTTCG	GACGGTTGGT	TTTTTACGGGA	TTTGAGATGT	TTGTTTTTA	420
GATTGTTGTT	TTTTTACGGGA	TGGCGCGGTG	TTTGGGTTTT	AGATTGTTA	GATAGATTAT	480
TTTTTGATGG	AGAGGGGATT	GTTCGCGCG	TTTCGGACGT	TTCGGGTTTT	GAGTTGCGGG	540
ATTTGTTTAT	CGGGCGCGAT	TTTTAGTAG	GTTCGCGTT	TTGTTTTTG	GTAAGTATCG	600
						639

ANGABEN ZU SEQ ID-NO:31:

SEQUENZCHARAKTERISTIKA:  
 LÄNGE: 304 Basen  
 ART: Nukleinsäure  
 STRANGFORM: Einzelstrang  
 TOPOLOGIE: linear

ART DES MOLEKÜLS: chemisch vorbehandelte Genom-DNA

## SEQUENZBESCHREIBUNG: SEQ ID-NO:31:

AGTAGTAGTA GTATAGTCG GCGTTGGTTA GCGGTTTAGA GTCGGGGTAG GGAGGGTTT	60
CGTGCGGACG GGGGTTAACGA GGTCGTTTT AGTCGGGGT AGTCGGCTA TTCGAGTTG	120
GAGTCGGAG TTTGGAGGGG TGCCCCAGGTT TCGCGTTAG GGTCGGATCG GAGGGAGAGG	180
GAGAGCGGTG GTTTTTTTT CGGTTTCGT CGTTAGTCGA TCGGGGCGTT GGGCAGGGCGT	240
CGCGGGAGTC GTAGTTTTT TCAGGGGGCG GATTGGTTT CGGCAGGGT ATTAGTTT	300
TGTT	304